

www.cell.com

www.cell.com



FOUR DECADES OF EXCITING BIOLOGY

40

SINCE 1974

Cell Press

CEO

Emilie Marcus

Vice President of Business Development

Joanne Tracy

Vice President of Operations

Keith Wollman

Publishing Directors

Peter Lee

Deborah Sweet

Editorial Director, Reviews Strategy

Katja Brose

Editorial Director, Content Development

Elena Porro

Director of Production

Meredith Adinolfi

Director of Marketing

Jonathan Atkinson

Media Relations Manager

Mary Beth O'Leary

Display Advertising

For advertising, sponsorship, and reprint opportunities, contact Jonathan Christison (jchristison@cell.com) or visit <http://advertisers.cell.com/contact-us>.

Recruitment Advertising

For recruitment advertising, visit <http://recruitmentads.cell.com/contact-us>.

©2014 Elsevier Inc. All rights reserved.

This journal and the individual contributions contained in it are protected under copyright by Elsevier Inc., and the following terms and conditions apply to their use:

Photocopying

Single photocopies of single articles may be made for personal use as allowed by national copyright laws. Permission of the Publisher and payment of a fee is required for all other photocopying, including multiple or systematic copying, copying for advertising or promotional purposes, resale, and all forms of document delivery. Special rates are available for educational institutions that wish to make photocopies for nonprofit educational classroom use.

Permissions

For information on how to seek permission, visit <http://www.elsevier.com/permissions> or call +44 1865 843830 (UK) / +1 215 239 3804 (US).

Derivative Works

Subscribers may reproduce tables of contents or prepare lists of articles including summaries for internal circulation within their institutions. Permission of the Publisher is required for resale or distribution outside the institution. Permission of the Publisher is required for all other derivative works, including compilations and translations (please consult <http://www.elsevier.com/permissions>).

Electronic Storage or Usage

Permission of the Publisher is required to store or use electronically any material contained in this journal, including any article or part of an article (please consult www.elsevier.com/permissions). Except as outlined above, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the Publisher.

Notice

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence, or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. Although all advertising material is expected to conform to ethical (medical) standards, inclusion in this publication does not constitute a guarantee or endorsement of the quality or value of such product or of the claims made of it by its manufacturer.

Reprints

Article reprints are available through Elsevier's reprint service; for information, contact Derrick Imasa (e-mail: d.imasa@elsevier.com; ph: +1 212 633 3874).

Subscription Orders and Inquiries

Mail, fax, or e-mail address changes to Elsevier Customer Service Americas, allowing 4–6 weeks for processing. Lost or damaged issues will be replaced, subject to availability, if Cell Press is notified within the claim period (US and airmail delivery: 3 months from issue date; surface delivery: 4 months from issue date). Periodical delivery in the US can take up to 3 weeks. Airmail delivery can take 2–4 weeks. Mailing address: Elsevier Customer Service Americas, Cell Press Journals, 3251 Riverport Lane, Maryland Heights, MO 63043, USA. Toll-free phone within USA/Canada: +1 866 314 2355; phone for outside US/Canada: +1 314 447 8880; fax: +1 314 447 8029; e-mail: subs@cell.com.

Funding Body Agreements and Policies

Elsevier has established agreements and developed policies to allow authors whose articles appear in journals published by Elsevier to comply with potential manuscript archiving requirements as specified as conditions of their grant awards. To learn more about existing agreements and policies, please visit <http://www.cell.com/cellpress/FundingBodyAgreements>.

Leading Edge

Cell Volume 159 Number 7, December 18, 2014

IN THIS ISSUE

EDITORIAL

1483 Punctuated Equilibria in Publishing

VOICES

1486 Guidance for Early-Career Scientists

PREVIEWS

- | | | |
|------|---|--|
| 1488 | Microexons Go Big | <i>L. Yang and L.-L. Chen</i> |
| 1490 | Super-Enhancer Transcription Converges on AID | <i>J. Alinikula and D.G. Schatz</i> |
| 1492 | Acetate Fuels the Cancer Engine | <i>C.A. Lyssiotis and L.C. Cantley</i> |
| 1494 | A New “Spin” on Recovery after Spinal Cord Injury | <i>A. Tedeschi and F. Bradke</i> |

REVIEW

- | | | |
|------|--|---|
| 1497 | Host Evasion and Exploitation Schemes of <i>Mycobacterium tuberculosis</i> | <i>C.J. Cambier, S. Falkow, and L. Ramakrishnan</i> |
|------|--|---|

SNAPSHOT

- | | | |
|------|--------------------------|--|
| 1712 | GPCR-Ligand Interactions | <i>E. Ghosh, K. Nidhi, and A.K. Shukla</i> |
|------|--------------------------|--|

(continued)

Articles

Cell Volume 159 Number 7, December 18, 2014

- 1511** A Highly Conserved Program of Neuronal Microexons Is Misregulated in Autistic Brains
M. Irimia, R.J. Weatheritt, J.D. Ellis, N.N. Parikshak, T. Gonatopoulos-Pournatzis, M. Babor, M. Quesnel-Vallières, J. Tapia, B. Raj, D. O'Hanlon, M. Barrios-Rodiles, M.J.E. Sternberg, S.P. Cordes, F.P. Roth, J.L. Wrana, D.H. Geschwind, and B.J. Blencowe
- 1524** B Cell Super-Enhancers and Regulatory Clusters Recruit AID Tumorigenic Activity
J. Qian, Q. Wang, M. Dose, N. Pruett, K.-R. Kieffer-Kwon, W. Resch, G. Liang, Z. Tang, E. Mathé, C. Benner, W. Dubois, S. Nelson, L. Vian, T.Y. Oliveira, M. Jankovic, O. Hakim, A. Gazumyan, R. Pavri, P. Awasthi, B. Song, G. Liu, L. Chen, S. Zhu, L. Feigenbaum, L. Staudt, C. Murre, Y. Ruan, D.F. Robbiani, Q. Pan-Hammarström, M.C. Nussenzweig, and R. Casellas
- 1538** Convergent Transcription at Intragenic Super-Enhancers Targets AID-Initiated Genomic Instability
F.-L. Meng, Z. Du, A. Federation, J. Hu, Q. Wang, K.-R. Kieffer-Kwon, R.M. Meyers, C. Amor, C.R. Wasserman, D. Neuberg, R. Casellas, M.C. Nussenzweig, J.E. Bradner, X.S. Liu, and F.W. Alt
- 1549** Apoptotic Caspases Suppress mtDNA-Induced STING-Mediated Type I IFN Production
M.J. White, K. McArthur, D. Metcalf, R.M. Lane, J.C. Cambier, M.J. Herold, M.F. van Delft, S. Bedoui, G. Lessene, M.E. Ritchie, D.C.S. Huang, and B.T. Kile
- 1563** Apoptotic Caspases Prevent the Induction of Type I Interferons by Mitochondrial DNA
A. Rongvaux, R. Jackson, C.C.D. Harman, T. Li, A.P. West, M.R. de Zoete, Y. Wu, B. Yordy, S.A. Lakhani, C.-Y. Kuan, T. Taniguchi, G.S. Shadel, Z.J. Chen, A. Iwasaki, and R.A. Flavell
- 1578** Dual Proteolytic Pathways Govern Glycolysis and Immune Competence
W. Lu, Y. Zhang, D.O. McDonald, H. Jing, B. Carroll, N. Robertson, Q. Zhang, H. Griffin, S. Sanderson, J.H. Lakey, N.V. Morgan, L.N. Reynard, L. Zheng, H.M. Murdock, S.E. Turvey, S.J. Hackett, T. Prestidge, J.M. Hall, A.J. Cant, H.F. Matthews, M.F.S. Koref, A.K. Simon, V.I. Korolchuk, M.J. Lenardo, S. Hambleton, and H.C. Su
- 1591** Acetate Dependence of Tumors
S.A. Comerford, Z. Huang, X. Du, Y. Wang, L. Cai, A.K. Witkiewicz, H. Walters, M.N. Tantawy, A. Fu, H.C. Manning, J.D. Horton, R.E. Hammer, S.L. McKnight, and B.P. Tu
- 1603** Acetate Is a Bioenergetic Substrate for Human Glioblastoma and Brain Metastases
T. Mashimo, K. Pichumani, V. Vemireddy, K.J. Hatanpaa, D.K. Singh, S. Sirasanagandla, S. Nannepaga, S.G. Piccirillo, Z. Kovacs, C. Foong, Z. Huang, S. Barnett, B.E. Mickey, R.J. DeBerardinis, B.P. Tu, E.A. Maher, and R.M. Bachoo

(continued)

- 1615** Sirtuin 4 Is a Lipoamidase Regulating Pyruvate Dehydrogenase Complex Activity
R.A. Mathias, T.M. Greco, A. Oberstein, H.G. Budayeva, R. Chakrabarti, E.A. Rowland, Y. Kang, T. Shenk, and I.M. Cristea
- 1626** Muscle Spindle Feedback Directs Locomotor Recovery and Circuit Reorganization after Spinal Cord Injury
A. Takeoka, I. Vollenweider, G. Courtine, and S. Arber
- 1640** The Cellular and Molecular Basis of Direction Selectivity of A δ -LTMRs
M. Rutlin, C.-Y. Ho, V.E. Abraira, C. Cassidy, C.J. Woodbury, and D.D. Ginty
- 1652** Detecting Envelope Stress by Monitoring β -Barrel Assembly
S.-H. Cho, J. Szewczyk, C. Pesavento, M. Zietek, M. Banzhaf, P. Roszczenko, A. Asmar, G. Laloux, A.-K. Hov, P. Leverrier, C. Van der Henst, D. Vertommen, A. Typas, and J.-F. Collet
- 1665** A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping
S.S.P. Rao, M.H. Huntley, N.C. Durand, E.K. Stamenova, I.D. Bochkov, J.T. Robinson, A.L. Sanborn, I. Machol, A.D. Omer, E.S. Lander, and E.L. Aiden
- 1681** X Chromosome Reactivation Dynamics Reveal Stages of Reprogramming to Pluripotency
V. Pasque, J. Tchieu, R. Karnik, M. Uyeda, A. Sadhu Dimashkie, D. Case, B. Papp, G. Bonora, S. Patel, R. Ho, R. Schmidt, R. McKee, T. Sado, T. Tada, A. Meissner, and K. Plath

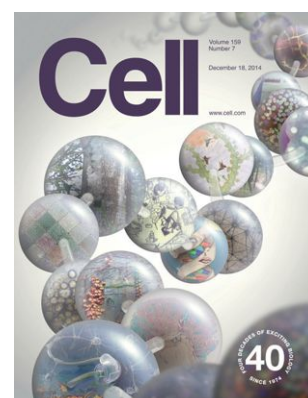
RESOURCE

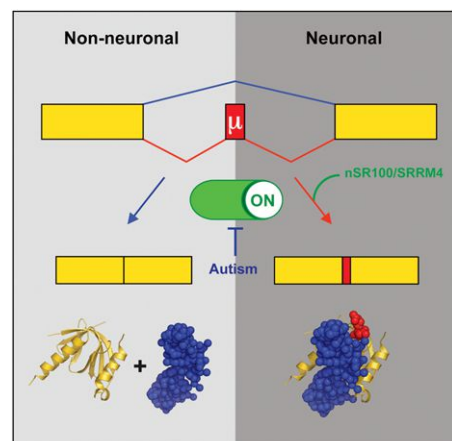
- 1698** High-Resolution Sequencing and Modeling Identifies Distinct Dynamic RNA Regulatory Strategies
M. Rabani, R. Raychowdhury, M. Jovanovic, M. Rooney, D.J. Stumpo, A. Pauli, N. Hacohen, A.F. Schier, P.J. Blackshear, N. Friedman, I. Amit, and A. Regev

ANNOUNCEMENTS

POSITIONS AVAILABLE

On the cover: *Cell* was founded 40 years ago as the “journal of exciting biology.” To close this celebratory year, we asked readers and authors to send us pictures that encapsulate their excitement about science. The image on the cover incorporates submissions received from established researchers and young scientists, spread throughout the world. It represents *Cell*’s commitment to serve the scientific community and the recognition that it is your passion and energy that truly make for “exciting biology.” An animated version of the cover, displaying additional images as well as thoughts from their contributors, is available online at <http://www.cell.com/40/cover>.





Misregulated Microexons in Autism

PAGE 1511

A myriad of alternative splicing events that provide important biological functions are not well understood. Irimia et al. discover and characterize hundreds of 3–27 nt neuronal-specific “microexons” in mammals. Relative to other classes of alternative splicing, neuronal microexons display the highest degrees of evolutionary conservation, frame preservation, and neuronal switch-like regulation and overlap with protein interaction domains. They are also frequently misregulated in autistic brains through a mechanism linked to the reduced expression of the neuronal-specific splicing regulator nSR100/SRRM4.

Super-Enhancers AID Cancer

PAGE 1524 and 1538

Activation-induced cytidine deaminase (AID) activity is required for antibody affinity maturation and class switch recombination in immunoglobulin genes. However, AID also acts on a set of off-target genes, generating translocations

and mutations that contribute to cancer. Qian et al. find that AID targets in the genome are interconnected in 3D networks that overlap with super-enhancer domains, revealing the role of the nuclear architecture and the B cell regulome in recruiting AID activity. Accordingly, Meng et al. find that the AID off-target activities are promoted by “convergent” sense/antisense transcription that emanates from super-enhancers within transcribed gene bodies. Together, these studies suggest that super-enhancers target oncogenes for translocations in cancer.

Pep-Tidying Up Amino Acid Pools

PAGE 1578

Lu et al. identify a new disease in humans caused by mutations in the enzyme Tripeptidyl peptidase II (TPPII), which manifests in the form of infections, autoimmunity, and neurodevelopmental delay. TPPII is required to maintain intracellular amino acid levels, and TPPII-deficient cells compensate by increasing lysosome number and proteolytic activity. The overabundant lysosomes derange cellular metabolism by consuming the key glycolytic enzyme hexokinase-2, which is what leads to impaired immune function.

Running on Acetate

PAGE 1591 and 1603

Two papers in this issue present evidence that acetate is a key metabolite for tumor cell sustenance. Comerford et al. use mouse models and analysis of human tumors to show that the acetyl-CoA synthetase enzyme, ACSS2, converts acetate into the key molecule acetyl-CoA in a manner that is required for tumor growth. Mashimo et al. employ NMR spectroscopy to determine that acetate is oxidized in vivo in both primary and metastatic tumors, concomitant with ACSS2 upregulation. Acetate dependence may represent a cancer Achilles’ heel.

A Sirtuin Surprise

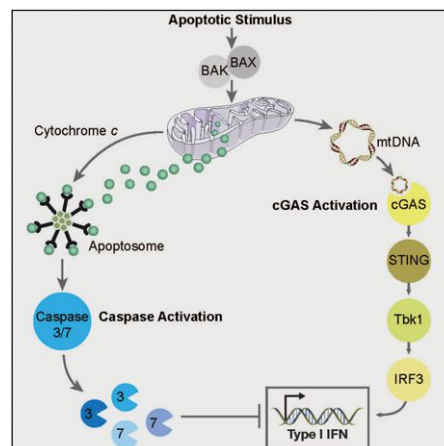
PAGE 1615

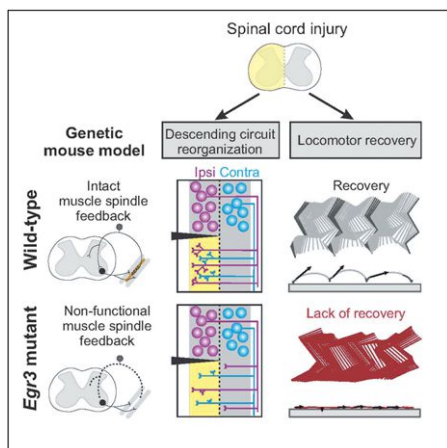
A study from Mathias et al. reveals two new aspects of cellular metabolism: a previously unknown enzymatic activity for a sirtuin protein and a noncanonical regulatory input into the activity of the pyruvate dehydrogenase complex (PDH). The mitochondrial SIRT4 functions as a lipoamidase that hydrolyzes lipoamide modifications on PDH (known previously to be regulated only by phosphorylation), leading to inhibition of its activity.

Caspases Prevent a STINGing Death

PAGE 1549 and 1563

Activated caspases are a hallmark of apoptosis induced by the intrinsic pathway, but they are dispensable for cell death and the apoptotic clearance in vivo. Now, White et al. and Rongvaux et al. independently find that mitochondrial events of apoptosis trigger the initiation of a cell-intrinsic immune response, mediated by the expression of type I IFNs. Proapoptotic caspases, activated simultaneously by mitochondria, are required to inhibit that response. In the absence of a functional caspases, dying cells behave as if virally infected, activating the cGAS/STING pathway to produce type I IFN, showing that the apoptotic caspase cascade functions to render apoptosis immunologically silent.





Bridging the Gap via Sensations

PAGE 1626

Spinal cord injuries alter motor function by disconnecting neural circuits above and below the lesion. Takeoka et al. now show that functional recovery upon injury relies on sensory feedback from muscle spindles, which facilitate the formation of detour circuits from the brainstem and spinal neurons and thereby bridge the injury.

Neurons Know Different Strokes

PAGE 1640

Direction-selective responses to stimuli are a key feature of several sensory systems, including the auditory and visual systems. Rutlin et al. show that a subset of cutaneous mechanosensory neurons, the A δ -LTMRs, is tuned to the direction of hair deflection. This property results from the developmental polarization of A δ -LTMR endings to the caudal side of hair follicles. BDNF emanating from hair follicle epithelial cells directs this process, rendering the ability of neurons to detect the direction of hair deflection.

larization of A δ -LTMR endings to the caudal side of hair follicles. BDNF emanating from hair follicle epithelial cells directs this process, rendering the ability of neurons to detect the direction of hair deflection.

Bacterial Stress Sensor

PAGE 1652

Detection of cell envelope integrity is essential for bacteria to rapidly respond to environmental stress. Cho et al. identify the lipoprotein RcsF as the sensor monitoring the functional integrity of the Bam machinery on the outer membrane. Bam binds RcsF and funnels it to the β barrel OmpA. Envelope stress interferes with this funneling ability of Bam, allowing the remaining RcsF to activate downstream signaling in response to environmental stress.

Hi-Res Hi-C and Well, Hello CTCF!

PAGE 1665

Rao et al. report a high-resolution Hi-C analysis of the human genome in multiple cell types that indicates a re-evaluation of chromatin domain organization and reveals unexpected insight into CTCF binding sites and inactive X chromosome topology. Distinct and conserved chromatin loops are defined as a domain subcompartments associated with different histone marks. Loop anchors typically occur at domain boundaries harboring CTCF sites arranged in a convergent orientation, with the asymmetric motifs “facing” one another.

Splicing RNA Thrill with DRILL

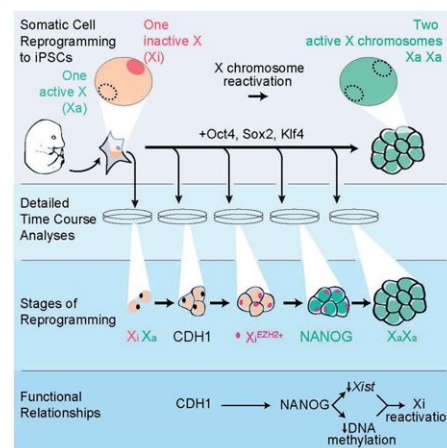
PAGE 1698

Rabani et al. present DRILL, a novel computational framework that uses RNA-seq data to discover transcriptional and posttranscriptional events that control dynamic changes in RNA transcript levels. The framework allows quantifying the level, editing sites, transcription, processing, and degradation rates of each transcript at a splice junction resolution and can be applied to coding and noncoding RNAs in different organisms.

X Inactivation in Reverse

PAGE 1681

Reprogramming somatic cells to iPSCs induces the reactivation of the inactive X chromosome (Xi). Tracking the epigenetic state of the Xi, Pasque et al. identify successive reprogramming stages and define requirements for Xi reactivation. The sequence by which Xi marks are reversed during reprogramming resembles the inverse order of developmental X inactivation for several marks but deviates from this chronology for marks associated with resistance to reprogramming. DNA methylation is particularly persistent, and its eventual removal appears to be independent of Tet enzymes.



Punctuated Equilibria in Publishing

Some species change little over long periods and then suddenly (at least on geological timescales) yield evolutionary innovations or diversify in creating new species. This herky-jerky relationship between time and change is known in evolutionary terms as punctuated equilibrium. Take, for instance, two extremes: the royal fern and the African cichlid fish. A time-traveling biologist visiting the world 180 million years ago would immediately recognize the former's leaves, roots, seeds, and shoots—even its nuclear structure and genome size resemble its modern-day counterpart (Bomfleur, B., et al. [2014]. *Science* 343, 1376–1377). Yet, if the same scientist visited the Great Rift Valley at intervals during the past 50 million years, she would witness periodic explosions of new types of cichlids, differing vastly in size, color, habitat, and behavior (the number of known species now surpasses 2,000 [Kocher, T.D. (2004), *Nat. Rev. Genet.* 5, 288–298]).

Disparities in rates of change are true of many evolving systems, biological or otherwise. The first scientific journals arose more than 300 years ago, and for much of the history of scientific publishing, the pattern of change has been more fern than fish. Yet, in the past 15 years, we have witnessed an explosion of new forms and innovations due to the selective pressures acting on the publishing ecosystem. These forces, to name a few, include new technologies, new business models, the introduction of government and funding body mandates, changes in peer review, the rise of big data science, and the growing trend toward interdisciplinary work. As a journal and as scientists, we now swim in depths more colorful, more exciting, and more competitive than ever before.

In marking the conclusion of *Cell*'s 40th anniversary celebration (<http://www.cell.com/40/home>), we look forward with excitement for what's next in science and for new opportunities in the world of scientific publishing. As with evolution, it's impossible to predict how publishing and science will change, but it is clear that, when we look back in 10 years from the vantage point of *Cell*'s 50th anniversary, we will see natural selection's indelible mark. Elements of this period of accelerated change are already evident in our recent “fossil record,” including the introduction of our policy in 2005 to open our archives so that all content from 1995 onward is accessible to everyone 12 months after publication; the introduction of our Article of the Future format for articles online, providing better and faster navigation through the article and incorporation of multimedia audio and video content; our introduction of a structured supplementary materials policy; and our recent focus on reproducibility and transparency in reporting ([http://www.cell.com/cell/fulltext/S0092-8674\(14\)01447-0](http://www.cell.com/cell/fulltext/S0092-8674(14)01447-0)). They are also evident in the changes in the science in our pages, as we move beyond our historical focus on mechanism to expand representation of big data science, human genetics, disease insights with clear therapeutic implications, and applied biology. *Cell* is excited to be a part of this transition and to help lead the way in ensuring that the new evolutionary equilibrium accelerates scientific advance and serves scientists

and society well. *Cell*'s “fitness” in this period of rapid evolution is empowered by our robust genomic integrity, our hybrid vigor from years of promoting the cross-pollination of ideas, and our agility (a.k.a., rapid generation times).

So what are some of the selective pressures currently shaping science and scientific publishing? Let's first look at the trajectories of change in science itself.

Era of Big Data

Perhaps the biggest shift has been from single-gene or protein-centric studies to increasingly common panoramic views of biology. Coming out of systems biology and the “-omics” revolution, many papers are now built on a foundation of large data sets. While the opportunity to see new patterns and insights from such panoramic views has clearly changed the way that we think about and understand many biological processes, it also creates obvious challenges for data management, accessibility, presentation, and peer review and further raises issues of where raw data should be stored and how access to it should be fostered. Tools for integrating data from multiple sources will become increasingly necessary to researchers, and there is an opportunity to make data sets as well as the papers in which they are published a focal point for community collaboration and discovery. Cell Press is currently working with our colleagues at Mendeley on building a data repository that will allow authors to easily host large data sets associated with their articles. To facilitate discovery, the repository will be searchable and integrated with community forums and collaboration groups. In addition, in recognition of the growing importance and excitement of big data science in biology and beyond, *Cell* is excited to welcome a new sister journal into the fold in 2015, with the launch of *Cell Systems*, a journal that showcases new breakthrough insights in biology at a systems level, new tools for systems-level analyses, and new applications of systems-level insight.

With the growth of global investment in biomedical research, we've also witnessed a shifting balance in centralized top-down versus experimenter-driven bottom-up research agendas. The development and output of large consortia such as ENCODE, TCGA, microbiome projects, and the global brain initiatives have put forth challenges about how to best communicate and disseminate the big-picture cumulative impact of geographically and temporally diverse collaborations. The current publishing system by necessity carves such large-scale consortia projects into individual articles published across multiple journals over multiple years, leaving the reader/user to “reintegrate” the individual pieces of the puzzle to appreciate the full impact of the original project vision. *Cell* is working with the leadership of some of these large-scale projects to see how we can evolve beyond traditional approaches to better communicate the full value and impact of these centralized initiatives and visions, with better linking and interaction between the puzzle pieces in the context of the bigger picture. As science evolves to become more collaborative, publishing must keep

pace with innovations to improve “collaboration between articles” as well.

Doing Science in a Changing World

Much of science today is truly interdisciplinary. Indeed, many of the most impactful studies use approaches that lie at the intersection of fields, a perfect example being the recent Nobel Prize in Chemistry for super-resolved fluorescent microscopy, a marriage of biology and physics. The confluence of basic biomedical research with fields such as engineering and clinical medicine represents the natural evolution of biology as a discipline. Reflecting this confluence, Cell Press recently collaborated with *The Lancet* to launch *EBioMedicine* (<http://www.ebiomedicine.com>), an open-access journal with a scope that spans the interface between biomedical research and clinical medicine. These exciting mergers will continue to drive discovery while simultaneously challenging how such multifaceted studies are effectively evaluated. It is an equally critical challenge to train the next generation of scientists to thrive in a world without borders between scientific disciplines.

As borders between disciplines are giving way, so too are the scientific borders between countries. Biomedical research, once the purview of select economies in the Americas, Europe, and Asia, is now a truly global effort. Though the focus, investment, and growth of Chinese biomedical research is widely recognized as representing the fastest rate of change, accelerated biomedical research is on the agenda of many other countries as well, including but not limited to India, Portugal, and Brazil. For *Cell*, this means that we have redoubled our efforts to connect with growing scientific communities worldwide, making sure that the doors are open for them to publish with us, that we build relationships with graduate students, postdocs, and scientific leaders in these countries, that we know who is doing breakthrough work where and on what, and that we build the international depth of our reviewer pool to capture the growing and sometimes unique expertise of different global communities. For example, Cell Press has been actively engaged with Chinese biomedical scientists since the early 2000s, and this year we made more than 40 editorial visits to institutes in China. In addition, we recently held our first Cell Symposium in Beijing, “Hallmarks of Cancer,” and we are continuing to build collaborations with leading institutions and scientists to ensure visibility of Chinese research on a global stage. In India, we host an annual week-long Distinguished Lecture tour with a leading international scientist in partnership with TnQ, an India-based publishing services and technology company. And finally, *Cell* editors travel to more than 60 conferences a year worldwide (for Cell Press as a whole, this number jumps closer to 300) to meet and engage with scientists as authors, reviewers, and readers and to hear exciting science.

Innovations in Scientific Communication

From scientific research to scientific communication, we are energized to continue to explore how best to effectively communicate exciting science to a broad audience. We frequently hear about the problem of information overload—scientists don’t have enough time to do all they need and want to do. Keeping up with the literature, staying current in their immediate

field, and finding time to be inspired by conceptual advances in other fields is a daunting task. Encouraging browsing and interdisciplinary thinking has always been and will continue to be a *raison d’être* for *Cell*, and in recent years we have introduced such features as research highlights and graphical abstracts to make it easier to understand at a glance the big conceptual message and contribution of each article. Nonetheless, the problem of information overload is unlikely to abate and, if anything, will become more critical. Platforms such as Mendeley help with customization and curation of scientific content, and in the future, we envision smart tool innovations that compile and summarize information relevant to a reader’s query or interests using big data analysis and peer recommendations (think Siri meets Amazon). Such advances in search tools will undoubtedly facilitate the type of open exploration and information grazing that will become increasingly central to the daily life of researchers.

Building on *Cell*’s “Article of the Future” project, we are committed to pushing the boundaries of how articles are presented online. This initiative, launched in 2010, led to the hierarchical presentation of text and figures and the now widely imitated sliding figure strip, allowing readers to more easily drill down through the layers of content based on their level of expertise and interest. The integration of research highlights, graphical abstracts, and other multimedia provides multiple mechanisms for conveying the core content of an article. There are now opportunities to revisit these principles with a fresh eye. As technology and reading habits continue to change, we are committed to staying up to date with what readers want and need to make the experience of perusing the contents of *Cell* even more engaging and stimulating. For example, we are currently piloting and getting feedback from readers on exciting prototypes for how technology can improve the way that data is effectively conveyed and absorbed in figures, so watch this space in 2015 for new developments in our “Article of the Future” initiative.

In recent years, there has been an expansion of business models competing in the publishing ecosystem, from author-pays open access (*Cell Reports* will soon have its third anniversary) to subscription, with many variants and hybrids in between, including funding from philanthropy and governments, each with their own pluses and minuses. Funding body mandates and the rise of open-access repositories to which many journals now permit posting of manuscript drafts (“green” open access) are other trends that will continue to shape how scientific information is stored and shared. Cell Press is proud to have been ahead of the curve in developing sustainable and innovative access models for high-quality content, and we are currently investigating new innovative approaches to ensure that every interested reader can access our content in a way that best serves his or her needs.

Now is also a time of experimentation in how manuscripts are evaluated prior to publication, including open peer review, single- and double-blinded peer review, collaborative peer review, and “post-publication” peer review. To this end, *Cell* is piloting a collaborative peer-review process in which reviewers are encouraged to comment on each other’s reports to consolidate the essential strengths and concerns prior to a final editorial decision. This pilot is still in the early stages for a subset

of manuscripts, but as we learn what works best to preserve rigor, excitement, and speed in peer review, we will look to roll out new tweaks to the system for all manuscripts. Whether these efforts will dramatically change the way that papers are vetted by expert appraisal is unclear, but we will continue to engage with the community as it gains experience with the merits and limitations of each approach. In support of reproducibility, *Cell*, together with NIH and other leading journals, has a renewed focus on experimental design and transparency in reporting, and to ensure adherence to ethical figure preparation guidelines, we have introduced a screening process for figures in accepted papers ([http://www.cell.com/cell/fulltext/S0092-8674\(14\)01447-0](http://www.cell.com/cell/fulltext/S0092-8674(14)01447-0)).

Evolving Measures of Merit

Let's face it, it's hard to find many fans of the impact factor, and if anything, dissatisfaction is growing. Yet, we as scientists like having data to analyze, and every professional ecosystem needs some at least somewhat objective means to separate signal from noise—to rate, rank, distinguish, and compare. As journals vary in the quality and rigor of their editorial processes, where a paper is published can and does tell you something about its relative significance and value, but a journal's "ranking" is not an absolute proxy for an article's quality, let alone that of its authors, and should never be used as a substitute for reading and assessing the actual science. Individual article citation counts also provide a means of assessing a study's impact, and some would argue that it is a better measure than a journal's average citation per article. But article-based metrics are also not without limitations as measures of quality. *Cell* has published many papers whose citations are well below our average and therefore would not score highly in an article-based citation count, but we consider them to be exciting, thought-provoking, rigorously supported conceptual advances that warrant a broad awareness, irrespective of whether they are in fields that are highly cited or not and even if it may take years for them to be fully integrated and built upon. For a fuller discussion of our thoughts on impact factors and measures of merit, see [http://www.cell.com/cell/fulltext/S0092-8674\(13\)00756-3](http://www.cell.com/cell/fulltext/S0092-8674(13)00756-3).

There are, of course, many ways in which a study's influence extends beyond citation. To reflect how an article is shaping the public scientific discourse, *Cell* and other Cell Press journals now provide Altmetrics for our articles, which tracks mentions in social media and news outlets and/or inclusion on such venues as Faculty of 1000. Although it is not yet clear what these various measures mean in terms of article quality and the work's ultimate contribution to advancing science, they are a real-time reflection of a paper's "buzz" and are clearly quite popular. Over the next decade, such tools will become more sophisticated in providing a well-rounded sense of a paper's impact and will become more personalizable to be tailored to a reader's particular interest.

How will these disparate forces shape *Cell* in the next decade? Or, in other words, are we fern or fish? The answer is both—our mission is to embody the best of each. Steadfast, like the royal fern, we will remain committed to publishing foundational and exciting research from across the broadest range of biology. Innovative, like the cichlids, we will adapt and pioneer new ways of reaching our readers, engaging our authors and reviewers, and communicating science globally. Over the course of *Cell*'s 40th anniversary year, we have had the opportunity to revisit many of the landmark achievements that have shaped the journal and its community of researchers (<http://www.cell.com/40/timeline>), to consider some of the emerging themes that will shape research in the next decade ([http://www.cell.com/cell/issue?pii=S0092-8674\(14\)X0007-3](http://www.cell.com/cell/issue?pii=S0092-8674(14)X0007-3)), and to highlight some of the rising stars whose creativity will fuel our future (<http://www.cell.com/40/under40>). From this reflection, we have a renewed appreciation that the journal succeeds as it adapts alongside its community, whether by expanding in scope or by challenging ways of thinking or presenting information. So, in essence, what happens to *Cell* is up to you. Our final cover of the year reflects this theme. As the "journal of exciting biology," we asked our readers to send us images that encapsulate what excites them most and have compiled them in a crowd-sourced cover with an animated version online (<http://www.cell.com/40/cover>). As we embark on our fifth decade together, our pledge to you is to nimbly respond to a changing scientific and publishing ecosystem. What we ask in return is simple: do what excites you.

The *Cell* editorial team

<http://dx.doi.org/10.1016/j.cell.2014.12.015>

Guidance for Early-Career Scientists

As part of *Cell*'s 40th anniversary celebration, we are spotlighting 40 principal investigators under the age of 40, and we asked each of them to give their personal advice to the upcoming generation of scientists. See the full profiles of all our "40 under 40" scientists and their responses to this and other questions at <http://www.cell.com/40/under40>.

A Tub Eventually Filling



Gloria Choi
Massachusetts Institute of Technology

As a new investigator, still feeling somewhat precarious and hoping for continued mentorship in my own career, it does feel odd giving out words of wisdom. I hope a description of my experience will suffice. I am a bit embarrassed to admit it, but I was very passive in the early decisions that have led to my career. Even in college, I majored in science because, as a nonnative English speaker, I thought I would not excel in the humanities. I never had a love-at-first-sight moment or a revelation that I am gifted (though I hope that I might be!), which led me to go to graduate school. Similarly, there was never one experiment that produced fantastic results to buoy me through the years of training. At the beginning, I just went to lab every day because that was the next step. But, as the small results piled up one after another, something changed and I began to truly love science. It was like a huge tub being filled with trickling water; you don't notice the meniscus moving, but it eventually fills up. One thing that was critical to this journey was that my advisors never let me lose sight of the finishing line made up of the big-picture questions we set out to answer. I am hugely indebted to them and their encouragement. So maybe the best advice I can give is to find great mentors who inspire you, and keep on working.

Enjoy the Journey



Jacob Hanna
Weizmann Institute of Science

"You are in science for the long run." I often have this conversation with PhD students and postdocs who are debating their choice of research track and next steps they should take. I remind them to "mentally" sit back, take their time, and enjoy the extended journey that is science and focus their energies on conducting more and more thoughtful experiments. Enjoy the benefits of doing a long PhD and learning as much as you can; enjoy a long postdoc where you can undertake challenging projects with the support of your host lab; don't hurry to finish and become an independent PI. Overall, these transitions are somewhat artificial, at least from a research perspective. In the end, scientific research is one continuous journey of endless learning and excitement, rather than jumps from position to position.

Expect the Unexpected



Rob Knight
HHMI, University of Colorado at Boulder

Improve your quantitative skills. Like astronomy two decades ago, biology is rapidly moving from an analog, data-limited science to a digital, data-rich science, and preparing yourself accordingly is essential. Maintain broad interests and don't over-specialize too early—you never know where the really exciting connections will come from. I certainly didn't predict that the graduate courses in community ecology and behavior that I was reluctantly forced to take in 1996–1997 would turn out to be essential to my research program 7–12 years later, but in retrospect, I am very glad that I had the relevant foundations in those areas.

Be Proud of Being a Biologist

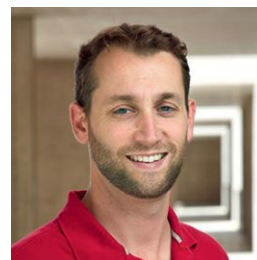
Aurelio Teleman
Deutsches Krebsforschungszentrum

One question that often comes up is whether a career in biology is the right choice. Indeed, biological research can be challenging—experiments don't always give clear results, and positive feedback comes infrequently. I was lucky to have the opportunity to experience the business world first hand and to notice that other jobs also have their own challenges if pursued at a high level. For instance, biological research comes with long-term stress; one might work on a project for several years without knowing if someone else in the world is also working on the same project with an associated risk of being scooped. In contrast, the business world is filled with short-term stress with deadlines that need to be met all the time. As another example, success in biology depends primarily on experimental results, whereas success in a company can largely depend on the opinions that other people have of you. I would keep this in mind when considering the challenges of biological research. The grass is not necessarily greener on the other side! Furthermore, biological research makes an important, useful, meaningful, and lasting contribution to our society and our civilization. That's more than what can be said of many other jobs. That makes me proud of being a biologist.

Be Open to Success

Jan Korbel
EMBL

If there is one thing that I have learned as a new PI, it is that obtaining feedback from others on your research ideas is as important as having a long-term vision for your group's research. Whether people approve of your plans or challenge them, feedback helps to refine and further sharpen your objectives. Openness, in this regard, can really pay off in the end. With this in mind, my advice for people who want to set up their own laboratory would be: do not make the mistake of not showing your job application, or manuscript, to peers before submitting. Don't let peer reviewers or group leader search panels be the first to read about your work. And although you don't have to follow every bit of advice you are given, it's helpful to be generally open to receiving, and providing, feedback.

Stay Positive

Bob Schmitz
University of Georgia

Try to ignore the negativity surrounding the poor funding environment and job prospects. This is an incredibly rewarding career that requires hard work. Things around me that I can't control, like current funding rates, often consume me, but my postdoctoral advisor Joseph Ecker always reminded me to keep my head down and do great things, good science will always be funded. The only thing I can control is the amount of time and effort I put into research. Let's be honest—as much stress as there is associated with a research career, it's a pretty fun job to be able to tinker around all day long in pursuit of the unknown. Find a scientific question that you're passionate about and begin the hunt.

Microexons Go Big

Li Yang^{1,*} and Ling-Ling Chen^{2,3,*}

¹Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology; CAS Center for Excellence in Brain Science

²State Key Laboratory of Molecular Biology, Shanghai Key Laboratory of Molecular Andrology, Institute of Biochemistry and Cell Biology Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

³School of Life Science and Technology, ShanghaiTech University, Shanghai, 200031, China

*Correspondence: liyang@picb.ac.cn (L.Y.), linglingchen@sibcb.ac.cn (L.-L.C.)

<http://dx.doi.org/10.1016/j.cell.2014.12.004>

Microexons are frequently underestimated in transcriptome analyses. Two studies published in *Cell* and *Genome Research* now independently report the identification of hundreds of microexons. Alternative splicing of some microexons is regulated by neuronal-specific RNA-binding proteins and modifies the function of proteins involved in neurogenesis, with misregulation linked to autism.

Nearly all human multi-exonic genes undergo alternative splicing (AS) to produce more than one mature mRNA, thus greatly expanding transcriptomic complexity and functional diversity (Nilsen and Graveley, 2010). Precise annotation of all AS events is essential for better understanding this repertoire under both physiological and pathological conditions, but the combinatorial aspect of this problem has been a major challenge. This problem is exacerbated in the case of microexons, exons less than 51 nt, which have often been overlooked because their short length makes them computationally difficult to identify. Microexons are thought to be unfavorable for splicing because they lack sufficient exonic splicing enhancers and they are so short that the splicing machinery cannot physically assemble at both the 3' and 5' splice sites (Black, 1991; Blencowe, 2000; Fairbrother et al., 2002). Individual studies in mammals reported an important role for microexons in the brain (Carlo et al., 2000; Zibetti et al., 2010), but the wider role of microexons and the rules governing their splicing have remained unclear. Now, two independent papers by Irimia et al. (2014) in this issue of *Cell* and by Li et al. (2015) in *Genome Research* uncover hundreds of highly conserved microexons from RNA-seq data sets across species, outline the

features regulating the inclusion of these microexons, and show that many of these impact neurogenesis and brain function.

To assess the contribution of microexons to the transcriptome, Irimia et al. (2014) develop a multi-module analysis pipeline to systematically define all neural-regulated AS patterns, especially microexons with very short lengths (3–15 nt), from more than 100 different human and mouse cell and tissue types. They show that the regulation of microexons is highly dynamic during neuronal differentiation (Figure 1). Strikingly, although microexons represent only 1% of AS observed, they constitute up to one-third

of all conserved neural-regulated AS between human and mouse. The inclusion in the final transcript of most identified neural microexons is regulated by a brain-specific factor, nSR100, which binds to intronic enhancer UGC motifs close to the 3' splice sites. Of particular interest, these microexons are enriched for lengths that are multiples of 3 nt and are thus highly likely to produce alternative protein isoforms if included or excluded from the final transcript. The authors further provide several lines of evidence, both computational and experimental, to demonstrate that inclusion of microexons can modulate the function of interaction domains of proteins involved in neurogenesis (Figure 1). Interestingly, misregulation of neural-specific AS microexons is observed in individuals with autism spectrum disorder (ASD).

Li et al. (2015) approach the issue in a different way. They treat microexons as insertions between annotated splice junctions to retrieve a set of microexons that are shorter than 51 nt, including both constitutively spliced (CS) and AS microexons, from more than 900 human and mouse samples, nearly half of which were from brain tissues. The authors find that AS microexons are evolutionarily conserved and exhibit tissue-specific inclusion. Furthermore, they show that AS of

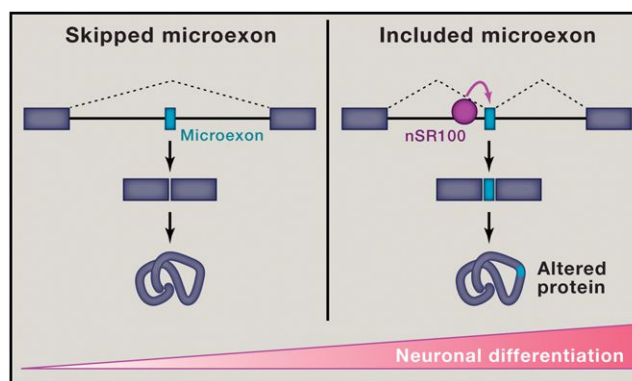


Figure 1. Alternatively Spliced Microexons Are Often Included during Neurogenesis along with the Expression of Neural-Specific Splicing Factors

Neural-specific microexons generally possess weak genomic features for splicing, such as unfavorable 3' splice sites, which leads to skipping at early stages of neuronal differentiation (left). During neurogenesis, some neuronal-specific splicing regulators, such as nSR100, become highly expressed and are recruited to intronic splicing enhancer regions near suboptimal 3' splice sites to promote inclusion of microexons. Inclusion leads to the addition of small numbers of amino acid residues in their protein products, which can alter protein-protein interactions (right).

microexons mediated by specific RNA-binding proteins (RBPs), such as RBFOX and PTBP1, may alter protein sequences, thus leading to changes in protein-protein interactions.

In contrast to CS microexons that possess strong *cis* elements to enhance splicing (Li et al., 2015), AS microexons require additional interactions with the splicing machinery, which are usually enhanced by RBPs. Interestingly, many brain-specific microexons might be regulated by a single RBP acting as a master splicing regulator. For instance, Irimia et al. (2014) demonstrate that neural-specific factor nSR100 promotes the AS of very short microexons during neurogenesis (Figure 1), and Li et al. (2015) confirm that most brain-specific microexons are enhanced by tissue-specific RBFOX proteins. It is noteworthy that RBFOX1-dependent AS has been implicated in ASD (Voineagu et al., 2011). These two new reports focus on different *trans*-acting factors, and questions remain, including the extent of the overlap of these two data sets and whether these distinct splicing regulators act independently or in concert to regulate AS of microexons in the brain. In addition, it will be of interest to identify other RBPs or master RBPs that can regulate AS microexons in different tissues or under different conditions.

Beyond the discovery of these surprising rules for the regulation of microexon

splicing, it is of particular significance that these two studies (Irimia et al., 2014; Li et al., 2015) demonstrate that alternative inclusion of microexons generates proteins with altered functions in neurogenesis. Whereas many microexons introduce short stretches of amino acids that alter protein-protein interactions, others may introduce novel charged regions or new platforms for post-translational modification. Not all lead necessarily to changes on the protein surface—one might envision that some microexon AS results in subtle alterations in protein folding or catalytic function. Furthermore, microexons can change the properties of the mRNA, altering its structure, stability, or subcellular location. Given the myriad ways that microexons can exert their influence, it is likely that they may have tissue-specific functions in other organs, and their mis-regulation may correlate with disease, as was observed for neuronal-specific microexons and ASD (Irimia et al., 2014). Also, as neural-specific microexon splicing is highly conserved during evolution at both the levels of genomic sequence and tissue-specific inclusion pattern (Irimia et al., 2014; Li et al., 2015), it will be of interest to study how selection acts on microexons. Together, these reports of the identification and impact of microexons demonstrate the feasibility of computationally probing transcriptome for previously hidden information and

begin to outline the mechanisms used by the cell to achieve the rich complexity of protein-protein interactions that govern tissue-specific processes.

ACKNOWLEDGMENTS

L.Y. is supported by 31271390 from NSFC. L.-L.C. is supported by XDA01010206 from CAS and 31322018 from NSFC.

REFERENCES

- Black, D.L. (1991). *Genes Dev.* 5, 389–402.
- Blencowe, B.J. (2000). *Trends Biochem. Sci.* 25, 106–110.
- Carlo, T., Sierra, R., and Berget, S.M. (2000). *Mol. Cell. Biol.* 20, 3988–3995.
- Fairbrother, W.G., Yeh, R.F., Sharp, P.A., and Burge, C.B. (2002). *Science* 297, 1007–1013.
- Irimia, M., Weatheritt, R.J., Ellis, J., Parikshak, N.N., Gonatopoulos-Pournatzis, T., Babor, M., Quesnel-Vallières, M., Tapia, J., Raj, B., O'Hanlon, D., et al. (2014). *Cell* 159, this issue, 1511–1523.
- Li, Y.L., Sanchez-Pulido, L., Haerty, W., and Ponting, C.P. (2015). *Genome Res.* Published online December 18, 2014. <http://dx.doi.org/10.1101/gr.181990.114>.
- Nilsen, T.W., and Graveley, B.R. (2010). *Nature* 463, 457–463.
- Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). *Nature* 474, 380–384.
- Zibetti, C., Adamo, A., Binda, C., Forneris, F., Toffolo, E., Verpelli, C., Ginelli, E., Mattevi, A., Sala, C., and Battaglioli, E. (2010). *J. Neurosci.* 30, 2521–2532.

Super-Enhancer Transcription Converges on AID

Jukka Alinikula¹ and David G. Schatz^{1,*}

¹Department of Immunobiology, Yale University School of Medicine and Howard Hughes Medical Institute, 300 Cedar Street, Box 208011, New Haven, CT 06520-8011, USA

*Correspondence: david.schatz@yale.edu

<http://dx.doi.org/10.1016/j.cell.2014.12.007>

AID mis-targeting is poorly understood but contributes significantly to B cell genome instability. Two new papers in *Cell* reveal that AID mistargeting occurs primarily in gene bodies within a nuclear microenvironment characterized by high levels of transcriptional activity, interconnected transcriptional regulatory elements, and overlapping sense and antisense (convergent) transcription.

During their development and function, B lymphocytes face a daunting series of challenges to genome integrity. Early in development, they must withstand multiple DNA double-strand breaks made by the RAG1/RAG2 endonuclease during assembly of immunoglobulin (Ig) heavy- and light-chain genes. Subsequently, activated B cells must deal with a bevy of mutations and DNA strand breaks triggered by the activation-induced deaminase (AID) during the processes of somatic hypermutation (SHM) and class switch recombination (CSR). Mistakes made during these reactions, including the erroneous targeting of non-Ig genes by RAG and AID, are the cause of many of the mutations and chromosomal aberrations found in B cell malignancies (Alt et al., 2013; Nussenzweig and Nussenzweig, 2010). Preventing mis-targeting of AID is a particular challenge because, unlike RAG, AID has no DNA-binding motif demarcating its appropriate target sites. It has proven difficult to explain why AID targets certain non-Ig genes, but not others—an issue of considerable importance because many erroneous AID targets are key B lineage regulators and potent proto-oncogenes. Two papers in this issue of *Cell* (Meng et al., 2014; Qian et al., 2014) take a major step forward in unraveling this mystery by linking AID mis-targeting to the process of convergent transcription within domains of highly interconnected transcriptional regulatory elements.

AID initiates SHM and CSR by deaminating cytosine residues in single-stranded DNA to yield uracil bases. The resulting U:G mismatches are processed into mutations (for SHM) or double-stranded breaks (for CSR) via DNA repair

pathways involving general base excision repair factors, mismatch repair factors, and error-prone DNA polymerases (Di Noia and Neuberger, 2007). AID targets are invariably transcribed, and AID interacts with a number of components of the transcription machinery, including RNA polymerase II (Pol II), the Pol II stalling factor Spt5, the single-strand DNA-binding complex RPA, and the RNA exosome. Such factors, acting in the context of stalled Pol II, are thought to recruit AID and create the single-stranded DNA substrate required for its action (Keim et al., 2013). Transcription per se, however, does not provide a ready explanation for why only certain transcribed non-Ig genes are targeted by AID or why Ig genes sustain mutations due to SHM at far higher levels than non-Ig genes (Storb, 2014). Attention has therefore focused on a central role for Pol II stalling, with CSR target regions (switch regions) providing an example of DNA sequences that favor the accumulation and stalling of Pol II and deamination by AID (Keim et al., 2013). It has remained a major challenge to understand the targeting preferences of AID elsewhere in the genome and to determine whether and how Pol II stalling might be involved.

Qian et al. and Meng et al. identified AID off-target DNA double-strand break (DSB) sites in the genome of activated B cells and intersected these data with an extensive array of epigenetic, nuclear architecture, and transcriptional data sets. Remarkably, most of the AID off-targets were found to lie within super-enhancers, large arrays of enhancers that accumulate high levels of activating histone marks, transcription factors, and components of the transcriptional machinery (Whyte

et al., 2013). Most of the DSB sites were found to lie within the region of overlap between a super-enhancer and the body of an active gene, but a small fraction fell in extragenic enhancers, which were themselves invariably transcribed. Qian et al. also found that the vast majority of AID-initiated lesions occurred near transcription start sites that were linked by long-distance interactions with multiple other promoters and enhancers, forming a “regulatory cluster.” Not all regulatory clusters or super-enhancers contained an AID off-target site, but those that did tended to be particularly large and have more “connectivity” (more linked promoters and enhancers) (Qian et al., 2014). Hence, the off-target activity of AID occurs preferentially in a particular nuclear microenvironment consisting of transcriptionally active, topologically highly interconnected, super-enhancer domains.

Why were only certain genes in this permissive nuclear microenvironment targeted by AID, and why did AID attack those genes in specific locations? Using very deep global run-on sequencing (GRO-seq) data, Meng et al. provide a remarkable answer to these questions: off-target AID-mediated DNA breaks almost invariably localized to sites of overlapping sense and antisense transcription, referred to as convergent transcription. Stronger super-enhancers and higher levels of convergent transcription correlated well with higher levels of AID-mediated DSBs. Together, the data of Meng et al. and Qian et al. lead to a model (Figure 1) in which AID-vulnerable sites in the genome are defined by the intersection of: (1) strong transcriptional activity (super-enhancers); (2) multiple

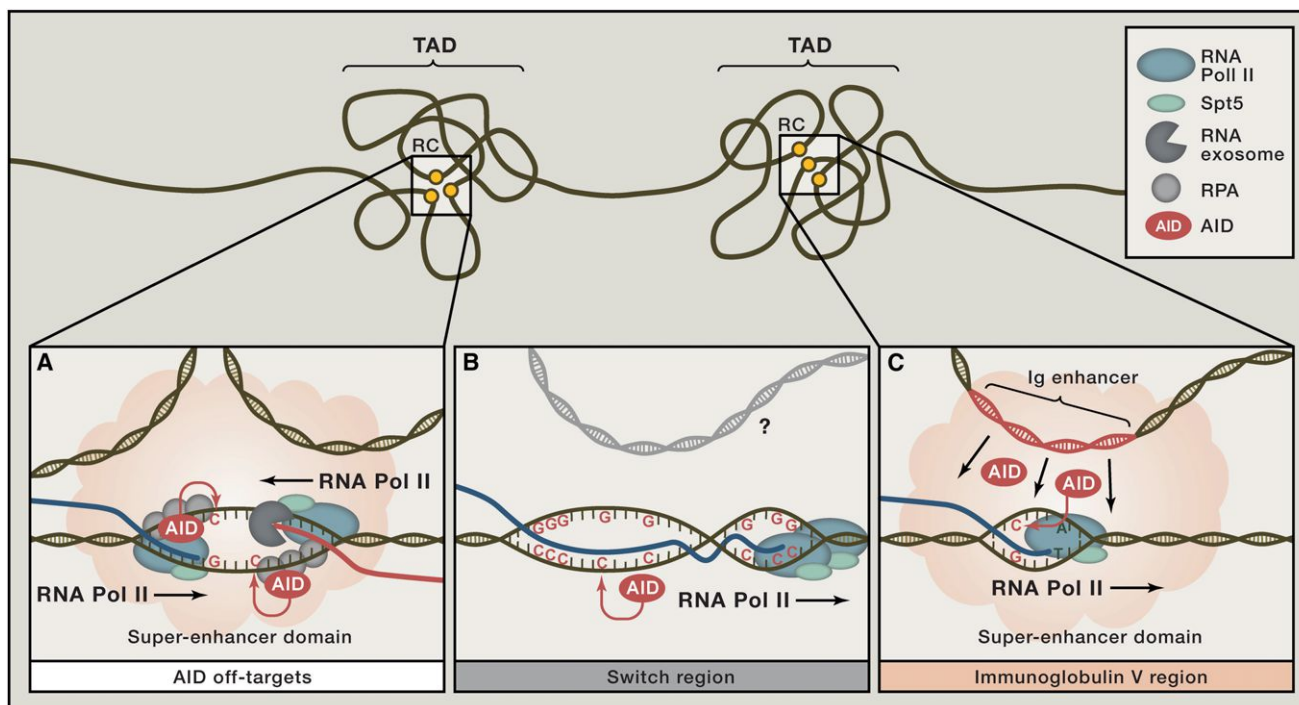


Figure 1. Targeting of AID Activity in the Genome by Super-Enhancers

(A) AID acts preferentially on non-immunoglobulin targets (“off-targets”) located in highly transcribed super-enhancer domains containing extensive looping between promoters and enhancers (regulatory clusters, RC), located within one topologically associated domain (TAD). Within these regions, AID acts at areas of convergent transcription where polymerases proceed in both directions generating sense- and anti-sense transcripts. These opposing polymerases collide and stall and, with the help of factors such as Spt5, RPA, and the RNA exosome complex, recruit AID and make single-stranded DNA substrates available for AID. (B) In immunoglobulin switch regions, special DNA properties (a G-rich nontranscribed strand and R loops) cause RNA polymerase II (Pol II) to stall and, together with multiple protein factors, to recruit AID and ensure single-stranded DNA accessibility. Whether super-enhancer features or convergent transcription further contribute to the targeting of AID activity to switch regions is unknown. (C) Targeting of AID and somatic hypermutation to immunoglobulin V regions are strongly enhanced by immunoglobulin enhancers by an unknown mechanism. Because mutations in germinal center B cells accumulate much more frequently in IgV region than in any AID off-target region, different/additional mechanisms are likely to be involved. A role for convergent transcription has not been ruled out.

interconnected transcriptional regulatory elements (regulatory clusters); and (3) strong convergent transcription, in which normal sense transcription of the gene overlaps with super-enhancer-derived antisense eRNA transcription. As Meng et al. point out, two RNA polymerases proceeding in opposite directions can collide and stall, thereby providing a favorable environment for the action of AID (Figure 1). Importantly, Qian et al. and Meng et al. extended their findings to human lymphomas, mouse germinal center B cells, and even mouse embryo fibroblasts. To a large extent, AID susceptibility tracked closely with the shifting landscapes of convergent transcription and super-enhancers. The one exception was AID-mediated deamination events detected as point mutations in repair-deficient, hypermutating B cells, where the correlations weakened somewhat, particularly for genes targeted at very low levels

by AID (only one-third of which displayed convergent transcription [Meng et al., 2014]). AID might therefore act at a low frequency outside of the permissive nuclear microenvironment defined by Qian et al. and Meng et al., a significant issue given the large number of such potential targets.

Substantial mechanistic questions and puzzles remain for those hoping to understand the targeting and mis-targeting of AID. The Pol II collision model is very attractive but now needs to be rigorously tested. How the clustering of regulatory elements contributes to AID action is not known; might this relate to the finding that Ig enhancers work together to target SHM (Buerstedde et al., 2014)? Not all sites of convergent transcription within super-enhancers are targeted by AID, suggesting that additional mechanisms might be layered on top of those uncovered by Meng et al.

and Qian et al. Although the Ig loci are found within super-enhancer regulatory clusters (Qian et al., 2014), it is not known whether convergent transcription contributes to the preferential targeting of AID to Ig variable regions (Meng et al., 2014). Finally, a recent study linked another unexpected transcriptional phenomenon—*divergent* antisense transcription upstream of transcription start sites—to the mis-targeting of AID (Pefanis et al., 2014). There appears to be much still to learn about the relationship between the antics of Pol II and those of AID.

REFERENCES

- Alt, F.W., Zhang, Y., Meng, F.L., Guo, C., and Schwer, B. (2013). *Cell* 152, 417–429.
- Buerstedde, J.-M., Alinikula, J., Arakawa, H., McDonald, J.J., and Schatz, D.G. (2014). *PLoS Biol.* 12, e1001831.

- Di Noia, J.M., and Neuberger, M.S. (2007). *Annu. Rev. Biochem.* 76, 1–22.
- Keim, C., Kazadi, D., Rothschild, G., and Basu, U. (2013). *Genes Dev.* 27, 1–17.
- Meng, F.-L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.-R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberger, D., et al. (2014). *Cell* 159, this issue, 1538–1548.
- Nussenzweig, A., and Nussenzweig, M.C. (2010). *Cell* 141, 27–38.
- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). *Nature* 514, 389–393.
- Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K.-R., Resch, W., Liang, G., Tang, Z., Mathé, E., Benner, C., et al. (2014). *Cell* 159, this issue, 1524–1537.
- Storb, U. (2014). *Adv. Immunol.* 122, 253–277.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). *Cell* 153, 307–319.

Acetate Fuels the Cancer Engine

Costas A. Lyssiotis^{1,*} and Lewis C. Cantley^{1,*}

¹Meyer Cancer Center, Department of Medicine, Weill Cornell Medical College, New York City, NY 10065, USA

*Correspondence: col2007@med.cornell.edu (C.A.L.), lcantley@med.cornell.edu (L.C.C.)

<http://dx.doi.org/10.1016/j.cell.2014.12.009>

Cancer cells have distinctive nutrient demands to fuel growth and proliferation, including the disproportionate use of glucose, glutamine, and fatty acids. Comerford et al. and Mashimo et al. now demonstrate that several types of cancer are avid consumers of acetate, which facilitates macromolecular biosynthesis and histone modification.

Metabolic pathways in cancer cells are programmed to facilitate survival and proliferation in the nonnative microenvironment of a tumor. This involves changes in both the way extracellular nutrients are captured and how they are metabolized. Historically, research efforts have focused on the wiring of glucose metabolism, owing to the seminal observations of Warburg and to the dominant role glucose plays in many basic biosynthetic processes (Vander Heiden et al., 2009). The importance of other fuel sources, including glutamine, lipids, and protein, have received more recent attention upon realization that pathways governing their metabolism are often driven by oncogenes. In this issue of *Cell*, new studies from the McKnight and Tu (Comerford et al., 2014) and Maher and Bachoo labs (Mashimo et al., 2014) illustrate that a variety of cancers are also capable of capturing and metabolizing exogenous acetate and that this represents a metabolic adaptation that some tumors use to facilitate growth.

Acetate, when ligated to coenzyme A (acetyl-CoA), is among the most central and dynamic metabolites in intermediary metabolism (Figure 1). It can be gener-

ated by the oxidation of glucose, glutamine, or fatty acids; it is used to biosynthesize nucleotides, amino acids, and both principle components of the cell membrane in mammals (i.e., fatty acids and cholesterol); and it contributes to enzyme and gene regulation by reversibly adding to nonhistone protein and histone tails, respectively (Figure 1) (Kaelin and McKnight, 2013). Indeed, numerous studies have illustrated the fundamental roles that acetyl-CoA regulation plays in cell growth and proliferative processes (Wellen and Thompson, 2012). However, under oxygen limiting conditions, as are often seen in the microenvironment of a tumor, the ability of a cell to make acetyl-CoA is severely hampered. Intrigued by this conundrum, and based on the mechanisms by which yeast generate acetyl-CoA, Comerford et al. (2014) explored the functional relevance of the mammalian homologs of the yeast enzymes that generate acetyl-CoA. Mammals express three isoforms of short-chain acyl-CoA synthetases (ACSS) that convert acetate and coenzyme-A into acetyl-CoA by consuming ATP. Two of these are localized in mitochondria (ACSS1 and

ACSS3), and one can access both the nuclear and cytoplasmic space, ACSS2 (Watkins et al., 2007). Comerford et al. (2014) find that knockdown of ACSS2, but not the mitochondrial isoforms, dramatically impairs the incorporation of exogenously supplied acetate into lipids and histone protein. These results illustrate that proliferating mammalian cells, including cancer cells, can consume and contribute acetate carbon to the cellular pool of acetyl-CoA.

In a parallel study, Mashimo et al. (2014) similarly find that exogenous acetate is captured and metabolized, here by human cancer cells grown in the brain of mice. The authors examined acetate metabolism in this context based on an earlier observation that a significant proportion of carbon in the acetyl-CoA pool could not be accounted for by tracing glucose and glutamine metabolism (Marin-Valencia et al., 2012). By tracing acetate carbon, Mashimo et al. (2014) reveal that TCA cycle intermediates consist of as much as 50% acetate-derived carbon by mass. In contrast, non-tumor-bearing brain incorporates on the order of 10% acetate-derived carbon into TCA cycle intermediates. These

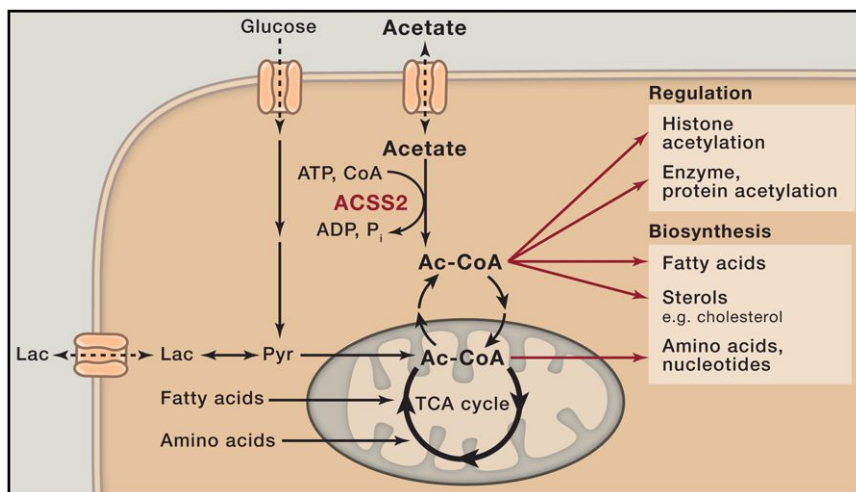


Figure 1. Acetyl-CoA Is a Central Node in Carbon Metabolism

Acetyl-CoA plays numerous roles in both regulatory and biosynthetic processes. It can be added in a posttranslational fashion to histone proteins to regulate gene expression or to other proteins and enzymes to dictate function or activity. Acetyl-CoA also serves as a principal building block for the generation of fatty acids, sterols, amino acids, and nucleotides. Acetyl-CoA is primarily generated in the mitochondria through catabolism of glucose, lipid, and amino acids. Mitochondrial acetyl-CoA is released into the cytoplasm by citrate export and breakdown. The cytoplasmic acetyl-CoA pool can also be filled by endogenous and exogenous acetate through reaction with coenzyme A by ACSS2.

observations are striking for several reasons, foremost because it illustrates that cancer cells are a sink for acetate and that acetate readily competes with glucose for generating TCA cycle intermediates, even though glucose is much more abundant. In these experiments, *in vivo* acetate levels are artificially raised to ~0.6 mM, and systemic glucose is maintained at ~5 mM. These results also provide validation of earlier, provocative work from the Bachoo and Maher labs (Marin-Valencia et al., 2012) that glutamine carbon is not actively oxidized in the TCA cycle in cells within the brain microenvironment (both brain cancer and metastases from other organ disease). Together, this argues that glucose and acetate metabolites are the dominant TCA cycle fuels in these cancers.

The independent but convergent findings from these two teams, which illustrate that acetate is readily captured and metabolized by cancer cells, prompted an exploration of the functional relevance and necessity of ACSS2 and acetate metabolism in cancer. To this end, Comerford et al. (2014) generate and cross an ACSS2 null mouse into two models of liver cancer. In both cases, the tumor burden is significantly blunted. Consistent with this finding,

high ACSS2 protein expression is observed in a subset of human triple-negative breast cancer samples, and this elevation correlates with poor survival. Similar correlations between ACSS2 elevation and poor outcome are obtained by Mashimo et al. (2014) in low-grade brain tumors (astrocytomas and oligodendrogliomas).

In both manuscripts, the authors illustrate that not only is ACSS2 expressed in tumors but that it is functional. Comerford et al. (2014) utilize radioactive carbon-labeled acetate, [¹³C]acetate, and PET imaging. They find that acetate avidity correlates well with ACSS2 expression in murine tumors. Tumors devoid of ACSS2 consume much less acetate, and tumors with high ACSS2 expression are acetate avid. Mashimo et al. (2014) utilize nonradioactive [¹³C]acetate and NMR to monitor *in vivo* acetate metabolism in tumors. The advantage of this technique is that the metabolism of acetate can be traced into downstream products. Initially, using orthotopic models with patient-derived material, they illustrate that acetate contributes a significant fraction of carbon to TCA cycle intermediates. Moreover, they show, using the identical technique in human patients, that brain tumors growing in human beings metabo-

lize acetate in a manner nearly identical to tumors grown orthotopically in the mouse brain.

Collectively, these results have several important therapeutic implications. First, they provide a clear demonstration that acetate is a metabolic fuel *in vivo* that is preferentially utilized by a subset of cancers. They also illustrate that this is mediated by ACSS2, whose expression correlates with tumor aggressiveness in the six different organ diseases analyzed. This suggests that acetate avidity and metabolism may be a general feature of many cancers. In contrast, normal cells appear unaffected by loss of ACSS2-mediated acetate metabolism, as ACSS2 null mice do not exhibit any overt phenotypic defects. Taken together, it is reasonable to conclude that acetate metabolism may represent an addiction of certain cancer cells. Furthermore, the results presented in these studies illustrate the clinical applicability of two different acetate-measuring technologies—i.e., [¹³C]acetate-PET and [¹³C]acetate-NMR. Such strategies could be used to identify patients likely to respond to an antimetabolism therapy and could also be used as markers of therapeutic response.

These findings beg the question as to how ACSS2 inhibition would affect an established tumor. The only experiment presented in this regard revealed that ACSS2 knockdown by short hairpin RNA (shRNA) in brain cancer cells grown in 3D culture resulted in cell death and an overall reduction in neurospheres. The astute reader may recognize that, under such conditions (growth in culture), very little acetate is present. More to this point, the concentration of serum acetate under physiological circumstances is on the order of 0.2 mM (Tollinger et al., 1979), raising the question of how a relatively low abundance molecule could contribute meaningfully to biomass in a rapidly proliferating cell. To address this, Comerford et al. (2014) put forth a persuasive argument for tumor cell evolution, relating it to that of bacterial cells evolving in a population, where seemingly minor advantages can ultimately have profound impacts. We put forth an additional, but not mutually exclusive, point of view based on three pieces of experimental data from these studies. First, exogenous acetate, by way of

acetyl-CoA, is found on histone proteins. Second, ACSS2 is readily observed in the nucleus of tumor cells, as evidenced by histological staining. And third, as noted above, ACSS2 knockdown in human patient tumor cell lines, grown in media devoid of acetate, is growth inhibitory. These findings suggest that the major role of ACSS2 is to capture acetate released from deacetylated proteins and to reincorporate that into the acetyl-CoA pool for epigenetic regulation. As Comerford et al. (2014) point out, the half-life of histone acetylation is on the order of minutes, and a considerable fraction of acetate could be produced in vivo by the turnover of histone acetylation. ACSS2 in the nucleus provides a rapid way to reconvert this acetate to acetyl-CoA for use in re-acetylating histones and thereby maintaining the epigenetic code. Although ACSS2 is not essential for this function in normal tissues, as evidenced by the viable ACSS2 knockout

mouse, it is possible that certain cancer cells require this function to maintain gene expression profiles optimized for rapid growth. Exogenous acetate, in this case, is treated equivalently to that generated by deacetylation. Regardless of the mechanism(s) by which cancer cells utilize acetate, the insights provided by these studies position acetate metabolism as a potentially exploitable vulnerability in cancer metabolism.

ACKNOWLEDGMENTS

C.A.L. is funded by a Pathway to Leadership grant from the Pancreatic Cancer Action Network and a Dale F. Frey Breakthrough award from the Damon Runyon Cancer Research Foundation. L.C.C. owns equity in, receives compensation from, and serves on the Board of Directors and Scientific Advisory Board of Agios Pharmaceuticals. Agios Pharmaceuticals is identifying metabolic pathways of cancer cells and is developing drugs to inhibit such enzymes in order to disrupt tumor cell growth and survival.

REFERENCES

- Comerford, S.A., Huang, Z., Du, X., Wang, Y., Cai, L., Witkiewicz, A.K., Walters, H., Tantawy, M.N., Fu, A., Manning, H.C., et al. (2014). *Cell* 159, this issue, 1591–1602.
- Kaelin, W.G., Jr., and McKnight, S.L. (2013). *Cell* 153, 56–69.
- Marin-Valencia, I., Yang, C., Mashimo, T., Cho, S., Baek, H., Yang, X.L., Rajagopalan, K.N., Maddie, M., Vemireddy, V., Zhao, Z., et al. (2012). *Cell Metab.* 15, 827–837.
- Mashimo, T., Pichumani, K., Vemireddy, V., Hatanpaa, K.J., Singh, D.K., Sirasanagandla, S., Nannepaga, S., Piccirillo, S.G., Kovacs, Z., Foong, C., et al. (2014). *Cell* 159, this issue, 1603–1614.
- Tollinger, C.D., Vreman, H.J., and Weiner, M.W. (1979). *Clin. Chem.* 25, 1787–1790.
- Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). *Science* 324, 1029–1033.
- Watkins, P.A., Maviguel, D., Jia, Z., and Pevsner, J. (2007). *J. Lipid Res.* 48, 2736–2750.
- Wellen, K.E., and Thompson, C.B. (2012). *Nat. Rev. Mol. Cell Biol.* 13, 270–276.

A New “Spin” on Recovery after Spinal Cord Injury

Andrea Tedeschi^{1,*} and Frank Bradke^{1,*}

¹Laboratory for Axon Growth and Regeneration, German Center for Neurodegenerative Diseases (DZNE), Ludwig-Erhard-Allee 2, 53175 Bonn, Germany

*Correspondence: andrea.tedeschi@dzne.de (A.T.), frank.bradke@dzne.de (F.B.)
<http://dx.doi.org/10.1016/j.cell.2014.12.014>

Functional recovery can occur after incomplete spinal cord injury. Takeoka et al. now report that such recovery relies on muscle spindle feedback that is necessary for neuronal circuit remodeling, suggesting novel targets to restore motor functions following spinal cord injuries.

Following incomplete lesions of the spinal cord, substantial recovery of sensory motor functions is observed (Curt et al., 2008; Martinez et al., 2012). Previous work has shown that such a recovery correlates with the formation of intraspinal circuits that bypass the injury (Bareyre et al., 2004; Courtine et al., 2008). Although sensory afferents are known to play a key role in the recovery process (Helgren and Goldberger, 1993), the sensory modality that allows the injured nervous

system to re-establish functional connections has remained elusive. In this issue, Takeoka et al. (2014) provide evidence for the role of muscle spindle feedback in promoting neuroplasticity and motor recovery following spinal cord injury (SCI).

Muscle spindles are sensory mechanoreceptors specialized for proprioception. They are located in skeletal muscles, and consist of several specialized intrafusal muscle fibers surrounded by a

capsule of connective tissue (Figure 1A). Muscle spindles are innervated by specialized motor and sensory axons. Deformation of intrafusal muscle fibers generates action potentials by activating stretch-sensitive ion channels expressed along the sensory axons that are coiled around the central part of the spindle. These axons connect to spinal motor neurons and different classes of interneurons that control muscle activity necessary for accurate body movements.

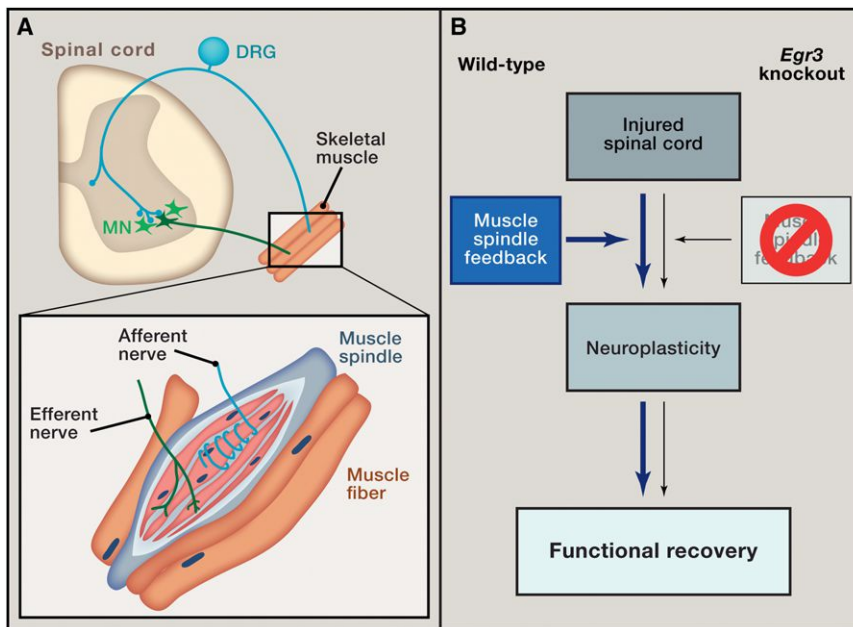


Figure 1. Promoting Neuroplasticity and Recovery after Spinal Cord Injury

(A) Muscle spindles are sensory organs located in skeletal muscles that receive innervation by specialized motor (efferent) and sensory (afferent) axons. Proprioceptive sensory axons originating from dorsal root ganglia (DRG) neurons spiral around the central region of the intrafusal fibers and respond to fiber stretch. The contractile regions of intrafusal fibers receive innervation by gamma motor neurons.

(B) Muscle spindle feedback promotes neuroplasticity and functional recovery after incomplete spinal cord injury. Absence of muscle spindle inputs in *Egr3* mutant mice results in impaired neuroplasticity and lack of recovery.

Previous studies demonstrated progressive postnatal degeneration of muscle spindles in mice lacking the zinc-finger transcription factor early growth response 3 (*Egr3*) (Tourtellotte and Milbrandt, 1998). While these mice are known to develop gait ataxia, resting tremors, and scoliosis (Tourtellotte and Milbrandt, 1998), the authors demonstrated with a sophisticated behavior/kinematic analysis combined with electromyogram recording that adult *Egr3* mutants have no major defects in walking when compared to wild-type mice. This provides a genetic entry point to study the contribution of muscle spindle feedback during motor recovery after SCI.

To study this process, the authors used thoracic lateral hemisection as a model of incomplete SCI. While control mice gradually recovered basic locomotor functions over time, *Egr3* mutants exhibited severe impairments on the ipsilateral side, providing evidence for a key contribution of muscle spindle feedback during the recovery phase and indicating that absence of physiological

inputs from muscle spindles may prevent engagement of spinal circuits. However, daily administration of monoaminergic receptor agonists, a pharmacological approach known to increase the activity of local spinal circuits (van den Brand et al., 2012), was not sufficient to promote locomotor recovery in *Egr3* mutants. This further supports the hypothesis that muscle spindle feedback directs motor recovery after incomplete SCI.

As recovery progresses, neuroplasticity occurs within the injured spinal cord. Several studies have shown that neuroplasticity can promote functional recovery in the absence of long-distance regeneration (Bareyre et al., 2004; Courtine et al., 2008). To investigate whether muscle spindle feedback contributes to this process, Takeoka et al. used state-of-the-art transsynaptic tracing techniques to determine changes in neuronal connectivity associated with recovery. To start off, the authors found no difference in the topographic organization of supraspinal pathways in wild-type versus

Egr3 mutant mice prior to injury. Several weeks after incomplete SCI, however, substantial reorganization of supraspinal pathways and the formation of midline-crossing detour circuits were found in wild-type but not in *Egr3* mutant mice (Figure 1B), thereby providing anatomical evidence that muscle spindle feedback promotes plasticity of neuronal circuits. It is important to note that motor skills that require fine control of body movements were also severely compromised in wild-type mice several weeks after injury, suggesting that muscle spindle feedback alone may not be sufficient to restore complex motor skills. This also highlights the importance of combinatorial strategies including the promotion of axon regeneration, task-specific rehabilitation, and/or electrical stimulation to refine connectivity of supraspinal pathways following SCI.

The work of Takeoka et al. clearly represents an important step forward for the field, once again underscoring the key role of muscle spindle feedback in directing locomotor recovery and circuit reorganization after injury (Figure 1B). However, a number of intriguing questions remain open. Earlier work demonstrated that absence of neurotrophin-3 (NT-3) in mutant spindles might be responsible for the lack of synaptic connectivity between sensory and motor neurons in *Egr3* mutants (Chen et al., 2002). Thus, an interesting experiment would be to test whether lack of substantial recovery in *Egr3* mutants could be restored by intramuscular viral delivery of NT-3. Furthermore, because *Egr3* mutant mice lack dual midline-crossing axons, it would be important to define whether a causal relationship exists between anatomical and functional outcomes. What happens after ablation of dual midline-crossing axons in wild-type mice? While this is an important experiment, it is worth mentioning that there are major difficulties hindering its execution. One major limitation is the absence of genetic markers to specifically select and manipulate the neurons from which midline-crossing axons originate. Future studies will be required to fully understand the molecular mechanism for muscle spindle feedback-mediated recovery after a variety of CNS trauma, including incomplete SCI.

ACKNOWLEDGMENTS

We thank Charlotte Coles and Wenjing Sun for critically reading. We apologize to authors whose relevant work we could not cite due to space limitations.

REFERENCES

- Bareyre, F.M., Kerschensteiner, M., Raineteau, O., Mettenleiter, T.C., Weinmann, O., and Schwab, M.E. (2004). *Nat. Neurosci.* 7, 269–277.
- Chen, H.H., Tourtellotte, W.G., and Frank, E. (2002). *J. Neurosci.* 22, 3512–3519.
- Courtine, G., Song, B., Roy, R.R., Zhong, H., Herrmann, J.E., Ao, Y., Qi, J., Edgerton, V.R., and Sofroniew, M.V. (2008). *Nat. Med.* 14, 69–74.
- Curt, A., Van Hedel, H.J., Klaus, D., and Dietz, V.; EM-SCI Study Group (2008). *J. Neurotrauma* 25, 677–685.
- Helgren, M.E., and Goldberger, M.E. (1993). *Exp. Neurol.* 123, 17–34.
- Martinez, M., Delivet-Mongrain, H., Leblond, H., and Rossignol, S. (2012). *J. Neurophysiol.* 108, 124–134.
- Takeoka, A., Vollenweider, I., Courtine, G., and Arber, S. (2014). *Cell* 159, this issue, 1626–1639.
- Tourtellotte, W.G., and Milbrandt, J. (1998). *Nat. Genet.* 20, 87–91.
- van den Brand, R., Heutschi, J., Barraud, Q., DiGiovanna, J., Bartholdi, K., Huerlimann, M., Friedli, L., Vollenweider, I., Moraud, E.M., Duis, S., et al. (2012). *Science* 336, 1182–1185.

Host Evasion and Exploitation Schemes of *Mycobacterium tuberculosis*

C.J. Cambier,¹ Stanley Falkow,² and Lalita Ramakrishnan^{1,3,*}

¹University of Washington, Seattle, WA 98195, USA

²Stanford University, Stanford, CA 94305, USA

³University of Cambridge, Cambridge CB2 1TN, UK

*Correspondence: lr404@cam.ac.uk

<http://dx.doi.org/10.1016/j.cell.2014.11.024>

Tuberculosis, an ancient disease of mankind, remains one of the major infectious causes of human death. We examine newly discovered facets of tuberculosis pathogenesis and explore the evolution of its causative organism *Mycobacterium tuberculosis* from soil dweller to human pathogen. *M. tuberculosis* has coevolved with the human host to evade and exploit host macrophages and other immune cells in multiple ways. Though the host can often clear infection, the organism can cause transmissible disease in enough individuals to sustain itself. Tuberculosis is a near-perfect paradigm of a host-pathogen relationship, and that may be the challenge to the development of new therapies for its eradication.

Introduction

Tuberculosis (TB) has afflicted humans for about 70,000 years and continues to take a huge toll on human life and health, with 8.6 and 1.3 million cases and deaths, respectively, in 2013 (Zumla et al., 2013; Comas et al., 2013). TB's timelessness in the face of significant human lifestyle changes over the millennia and the advances of modern medicine over the last century bespeak the agility and toughness of its causative pathogen *Mycobacterium tuberculosis*. *M. tuberculosis* may be the paradigm for human host-pathogen adaptation.

TB's notoriety as one of the great bacterial terrors of humanity alongside plague, typhus, cholera, typhoid, and diphtheria has led to descriptors such as the "great white plague" and "the captain of all those men of death." When compared to other major bacterial diseases, there are some interesting and potentially informative aspects of TB's pathogenesis. Human infection and disease is essential for the transmission and therefore the evolutionary survival of *M. tuberculosis*. This is in contrast to plague, which, despite its enormous impact on human history, is a zoonosis in which human disease is essentially an accident with no bearing on the pathogen's subsequent survival. The same could be said for many commensal pathogens, e.g., the pneumococcus, meningococcus, or the flesh-eating streptococci, in which human disease, though terrifying, is of marginal benefit to the long-term survival of the pathogen. Despite the inextricable connection between disease and transmission and thereby its survival, *M. tuberculosis* appears to lack the classical virulence factors that are the badges of honor of many of these pathogens. These include capsules to avoid phagocytosis, pili, or other adhesins for adherence to host tissues; flagella for motility; and enzymes and toxins to poison host cells. How does *M. tuberculosis* produce the disease so devastating to humans and so vital to the pathogen? The classical virulence factors of the mucosal commensal pathogens, many of which reside in the nasopharynx, are really colonization factors that,

in the right host, run amok to cause a disease that is of questionable benefit to the pathogens' evolutionary survival. These factors probably give the microbe a selective colonization advantage on mucosal surfaces, where bacterial competition is rife. Not surprisingly, vaccines against individual virulence factors—be it a capsule of the pneumococcus or a toxin—sometimes eradicate colonization along with disease. As an example, the diphtheria toxin that has been responsible for countless million deaths in the past is likely a colonization factor that allows *Corynebacterium diphtheria* to compete effectively with other mucosal bacteria to establish a privileged niche in the tonsils. The diphtheria vaccine that is directed against this single toxin has wiped out colonization together with disease. These represent the more recently evolved "crowd" diseases that emerged in the neolithic age associated with the development of agriculture and the domestication of animals. In contrast, *M. tuberculosis* is an ancient companion of man since before the neolithic age and its associated crowding (Comas et al., 2013), making TB a "heritage" disease for much of its history. We argue that *M. tuberculosis* and many of the other host-adapted mycobacteria have evolved a different strategy for insuring persistence in the host—they have honed their lifestyle to obviate the need for virulence (née colonization) factors like toxins and enzymes that break down anatomic barriers to outcompete other pathogens. Instead they use host macrophages to traverse host mucosal barriers to sterile sites deep in the body. As we see it, *M. tuberculosis*' dirty little secret is to be hydrophobic and to "fly" more efficiently in a tiny droplet to bypass the innate immune system. Rather than jostling with other pesky microbes, *M. tuberculosis* can deal just with its host, and we suggest that host immune evasion, modulation, and exploitation are the trump cards of the pathogenic mycobacteria. This recognition of *M. tuberculosis*' tactics brings a new understanding of host and pathogen biology that can potentially be parlayed into new therapies and interventions.

This Review will examine key new discoveries about TB pathogenesis against a backdrop of the natural history of infection and disease and its difficult treatment. We note that TB pathogenesis was last reviewed in *Cell* in 2001 (Glickman and Jacobs, 2001), a time that marked the derivation of a basic molecular genetic toolkit for *M. tuberculosis* and the postgenomic era of TB research being ushered in following the elucidation of its genome sequence (Cole et al., 1998). The earlier *Cell* Review highlighted the problem of lengthy drug treatment as a factor that made global eradication of TB difficult, described new insights into the strategies used by *M. tuberculosis* to persist in macrophages, and discussed newly identified lipid effectors in virulence. Since then, many additional mycobacterial genomes have been sequenced, enhancing our understanding of mycobacterial evolution. More sophisticated genetic approaches and new animal models have provided new and often surprising insights into how the pathogenic mycobacteria survive and replicate in macrophages and indeed orchestrate the formation of granulomas, macrophage aggregates, and exploit them for their expansion. We will discuss how the mechanisms used by mycobacteria to resist macrophages also render them drug tolerant, a finding that has potential therapeutic implications. Finally, we will discuss how mycobacteria paradoxically can benefit from an over-exuberant host immune response to increase their numbers further and be transmitted to a new, susceptible host. The significance of these discoveries may be most fully appreciated in the context of both mycobacterial evolution and host adaptation. Where appropriate or interesting, we will compare or contrast mycobacterial pathogenic strategies to those of other pathogens. Many of our insights and ideas have come using *Mycobacterium marinum*, a close genetic relative of *M. tuberculosis* (Stinear et al., 2008), which we have developed as a valid and tractable model for *M. tuberculosis* pathogenesis (Ramakrishnan, 2013; Tobin and Ramakrishnan, 2008). We will therefore use *M. marinum* as a stand-in for *M. tuberculosis*, as well as a comparator. We will organize our thoughts around the “pathogenic personality” of *M. tuberculosis* and its many facets as it goes through its pathogenic lifecycle—entry into the host, attainment of a unique niche, multiplication within, and exit from the host—all by avoiding, circumventing, or manipulating host defenses with a unique “pathogenic signature” (Falkow, 2008) (Figure 1).

Stealth Entry Affords Mycobacteria a Privileged Host Niche

The optimal niche for a host-adapted pathogen within a host is the environment in which the pathogen is readily able to replicate. To arrive at and replicate in this niche, pathogens must circumvent host defenses, (Falkow, 2006), which are in turn substantially influenced by the commensal microflora that abundantly populate our skin and mucosal surfaces. The host-commensal alliance that forms the barrier to pathogens has recently been reviewed in *Cell*, as part of this 40th anniversary series (Belkaid and Hand, 2014).

M. tuberculosis is known to initiate infection most efficiently in the lower lung through small aerosol droplets that contain only one to three bacteria, a constraint that makes it less contagious than respiratory pathogens, such as the group A streptococcus

and *Corynebacterium diphtheriae* that initiate infection in the nasopharynx to cause, respectively (1) strep throat and scarlet fever and (2) diphtheria, which are spread through large, wet droplets. The insight that small droplets are most likely responsible for transmitting TB comes from human epidemiological studies examining transmission from index cases in confined spaces (Bates et al., 1965; Houk, 1980). Corroborating the human studies are studies that track infection serially in rabbits, demonstrating that aerosol droplet size negatively correlates with infection burdens (Wells et al., 1948). When large aerosolized particles containing 10,000 bacteria were administered, they got stuck in the trachea and the rabbits got no or very little infection. In contrast, upon receiving small aerosols containing one to three bacteria that reached the alveolar spaces of the lung, the rabbits all got progressive infection.

A teleological explanation for why TB initiates in the lower lung at the cost of infectivity comes from the zebrafish larval model of TB (Cambier et al., 2014). The zebrafish larva is optically transparent so that infection with fluorescently labeled bacteria can be monitored in exquisite detail (Takaki et al., 2013). To examine the earliest interactions with the host, bacteria can be injected into the hindbrain ventricle, a microbiologically sterile neuroepithelium-lined cavity in which macrophages and neutrophils are not present normally but migrate with the expected specificity in response to the microinjection of specific chemokines or bacteria (Takaki et al., 2013; Yang et al., 2012). The zebrafish work suggests that pathogenic mycobacteria have developed strategies to avoid the microbicidal macrophages that are the default recruits to keep mucosal commensal pathogens at bay. These macrophages already primed to be microbicidal are recruited through Toll-like receptor (TLR)-mediated signaling that is activated by the so-called pathogen-activated molecular patterns (PAMPs) present on bacterial surfaces. In mouse and zebrafish macrophages, the TLR-induced microbicidal activity is from reactive nitrogen species produced by the action of inducible nitric oxide synthase (iNOS) (Cambier et al., 2014); different microbicidal effectors may be induced in human macrophages (Liu et al., 2006). Mycobacteria, including *M. tuberculosis* and *M. marinum*, are replete with PAMPs. Indeed, complete Freund's adjuvant that is used to prime immune responses is nothing but an oil emulsion containing dead *M. tuberculosis*. However, these mycobacteria express a surface lipid phthiocerol dimycocerosate (PDIM) that masks the PAMPs so that they are not “seen” by the host innate immune system. Concomitantly, they use a related surface lipid, phenolic glycolipid (PGL), to induce the macrophage chemokine CCL2 to recruit and infect macrophages that are growth-permissive for them. However, this strategy of using a masking lipid to avoid the microbicidal macrophages and a recruiting lipid to infect the permissive ones would be ineffective in the upper airway, an environment replete with an endless supply of TLR-stimulating commensal bacteria. On this battlefield, mycobacteria would be collateral damage caught in the crossfire; they would be killed by the microbicidal macrophages that are continually being recruited. Hence the need for the third component of their tripartite immune evasion strategy: small infection droplets that deliver them directly into the alveolar spaces of the lower lung, which harbors few, if any, commensals (Charlson et al., 2011) (Figure 2).

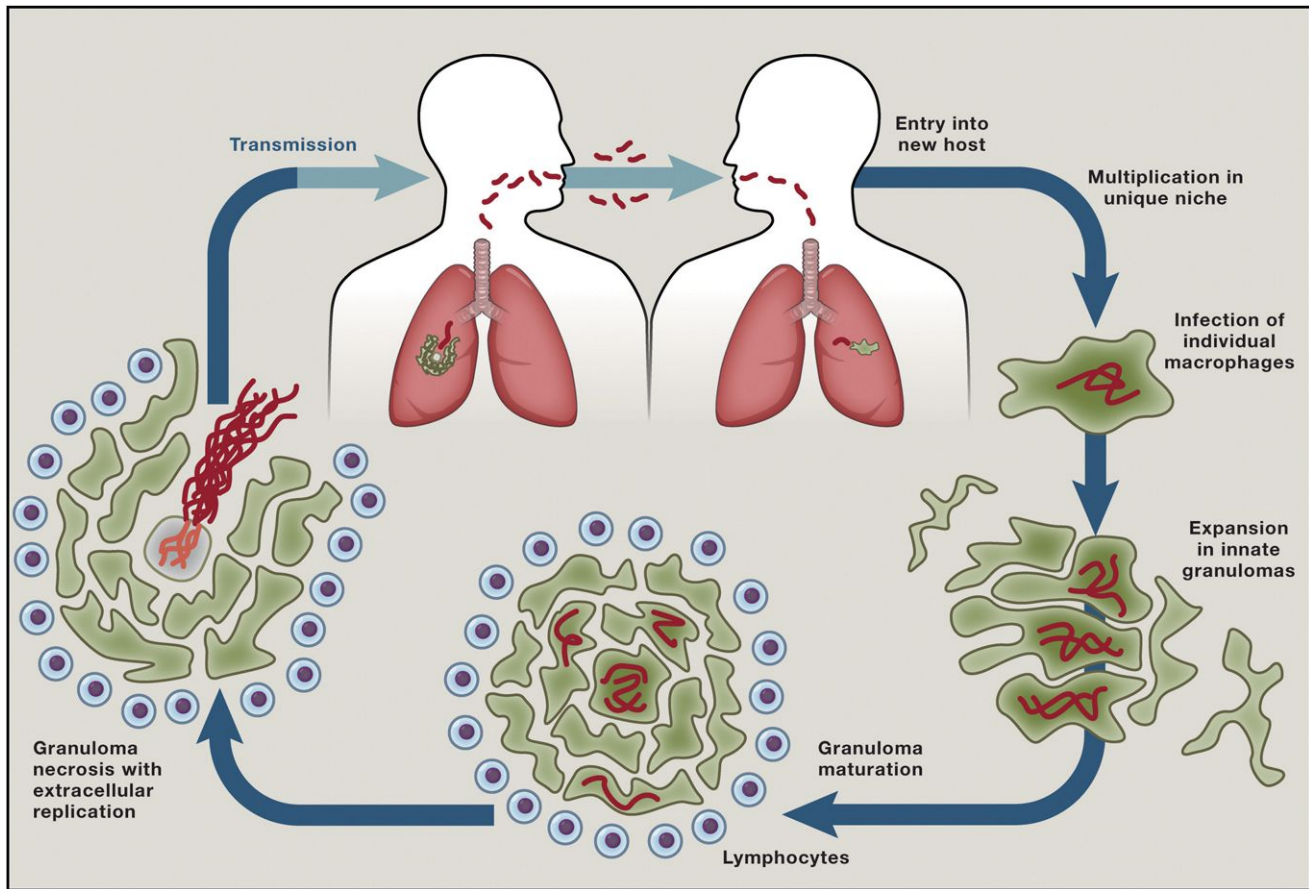


Figure 1. Pathogenic Life Cycle of *M. tuberculosis*

M. tuberculosis infection initiates when fine aerosol particles containing the bacteria coughed up by an individual with active disease are deposited in the lower lungs of a new host. The bacteria recruit macrophages to the surface of the lung, which become infected, and serve to transport the bacteria across the lung epithelium to deeper tissues. A new round of macrophage recruitment to the original infected macrophage is initiated, forming the granuloma, an organized aggregate of differentiated macrophages and other immune cells. The granuloma in its early stages expands infection by allowing bacteria to spread to the newly arriving macrophages. As adaptive immunity develops, the granuloma can restrict bacterial growth. However, under many circumstances, the infected granuloma macrophages can undergo necrosis, forming a necrotic core that supports bacterial growth and transmission to the next host.

There is a growing appreciation for a commensal-primed barrier immunity that pathogens must evade, tolerate, or interrupt. *Helicobacter pylori*, a commensal pathogen that famously causes gastric ulcers, is also a heritage pathogen and has adapted to survive in the stomach, where competition from commensals is minimal (Monack, 2013). *H. pylori* too has evolved to avoid detection via TLRs: its flagellin is not recognized by TLR5 (Gewirtz et al., 2004), and its lipopolysaccharide (LPS) has a lower affinity for TLR4 than that of other bacteria (Moran, 2007). Therefore, like *M. tuberculosis*, *H. pylori* has evolved to avoid proinflammatory host detection by initiating infection in anatomical locations in which commensal competition is minimal. Although *M. marinum* and *M. tuberculosis* have developed tactics to evade reactive nitrogen species, *Mycobacterium avium* that causes TB-like disease in birds appears to have evolved a strategy to tolerate and even benefit from them and, accordingly, does not express PDIM (Dhama et al., 2011; Dumarey et al., 1994; Gomes et al., 1999; Onwueme et al., 2005) (Figure 3). The case of host-adapted *Salmonella*,

another macrophage-dwelling class of pathogens, may be illustrative as well. *Salmonella* infects via the terminal ileum that is replete with colonizing bacteria. To facilitate its transit through the commensal-laden gut, *Salmonella* appears to first drive an inflammatory response that generates the reactive nitrogen species to which the commensals are sensitive, but it, like *M. avium*, is tolerant at least early during infection (Fang, 2004; Henard and Vázquez-Torres, 2011). Upon reaching the terminal ileum, the invading *Salmonella* enters into M cells, specialized cells of the follicle-associated epithelium, a region that is again relatively free from commensal competition (Jones et al., 1994), and invades underlying macrophages. This multi-pronged strategy to interrupt the commensal barrier so as to reach the M cells affords *Salmonella* access to the systemic phagocytes of the host (called the reticuloendothelial system). The common theme emerging from these scenarios is that host-adapted pathogens must develop strategies to circumvent the host-beneficial commensal-primed immune barrier in order to reach their replicative niche. We emphasize that, in turn,

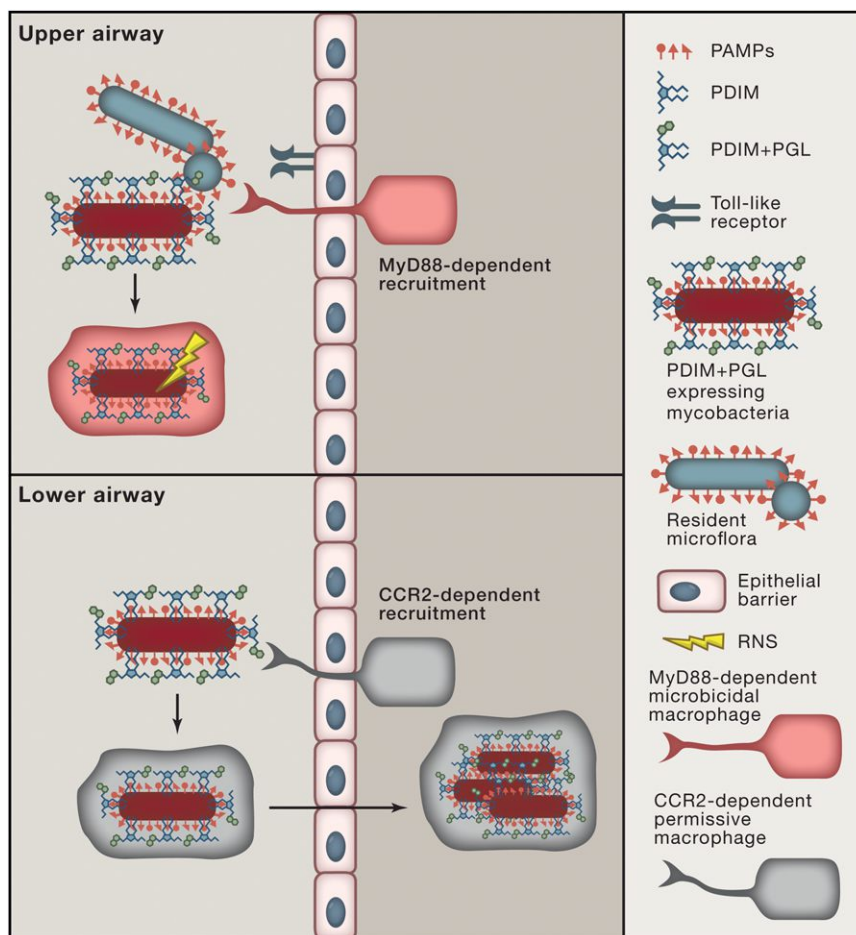


Figure 2. *M. tuberculosis* Evades Commensal Bacteria to Infect Its Host

M. tuberculosis avoids the recruitment of microbicidal macrophages to the site of infection by masking its PAMPs with the PDIM lipid. A related surface lipid PGL recruits permissive macrophages that can transport the bacteria into deeper tissues. However, the upper airways are colonized by resident microorganisms whose PAMPs recruit microbicidal macrophages. Therefore, this mycobacterial strategy to evade microbicidal macrophages is only effective if infection is initiated in the relatively sterile lower lung.

surface substantially increases the organism's chances of reaching its preferred replicative niche. These findings may also serve to explain human studies showing an association between TB susceptibility and the high expression of CCL2, PGL's host partner in recruiting permissive macrophages (Flores-Villanueva et al., 2005). Further, the finding that PGL increases virulence through enhanced infectivity provides an understanding of why PGL is present in *M. canettii* strains, ancestral to *M. tuberculosis*, as well as in *M. marinum*, the closest genetic relative of the *M. tuberculosis* complex (Onwueme et al., 2005), suggesting its integral role in the evolution of mycobacterial pathogenicity. As noted before, TB is generally thought to have infected humans for ~70,000 years, thus predating by

commensals exert a selective pressure that has shaped pathogen evolution.

Mycobacteria have to engage with the host to become phagocytosed by permissive macrophages by simultaneously using the surface lipid, PDIM, to dampen TLR signaling and by using the related surface lipid, PGL, to induce CCL2 signaling. This finding has additionally provided an understanding of the role of these lipids in virulence (Siegrist and Bertozzi, 2014). PDIM is expressed only by pathogenic mycobacteria, is absolutely required for virulence, and is present in all *M. tuberculosis* clinical isolates (Onwueme et al., 2005). Yet PDIM synthesis is metabolically costly so that it is readily lost in axenic culture (Kirksey et al., 2011). PGL, in contrast, is not absolutely required for virulence; it is not present in all clinical *M. tuberculosis* isolates (Reed et al., 2004). However, it is present in many of the W-Beijing strains that have predominated in outbreaks in North America, where TB is not prevalent. In the zebrafish, wherein low-dose infections can be examined longitudinally from the first instances of infection, PGL specifically increases infectivity of inocula of one to three bacteria (that mimic those of human infection) by enhancing the recruitment of mycobacterium-permissive macrophages. It is only in the context of examining the ability of the pathogen to establish infection, rather than sustain it, that the role of PGL is revealed. The presence of PGL on the bacterial

~60,000 years the neolithic demographic transition and its resultant crowding (Bos et al., 2014; Comas et al., 2013). Thus, PGL may have been an essential virulence determinant for most of its history. Perhaps, the greatly increased transmission opportunities arising from human crowding made it dispensable.

Multiplication within the Host—The Macrophage Niche

The strategy elaborated by *M. tuberculosis* to traverse host epithelial barriers within permissive macrophages is, of course, predicated upon its ability to survive within these highly evolved phagocytic host cells. Indeed, macrophages comprise the replicative niche for most of the lifecycle, not only of *M. tuberculosis* but of most other pathogenic mycobacteria (Figure 4). Accordingly, the ability to replicate in host cells is a defining feature of the pathogenic mycobacteria—be they human or animal pathogens—and reliably distinguishes them from their nonpathogenic soil-dwelling cousins like *Mycobacterium smegmatis* (Shepard, 1957) (Figure 3). A clue for how this ability to grow in host macrophages might have evolved comes from the remarkable finding that the ability of mycobacteria to replicate in macrophages tracks completely with their ability to grow in unicellular free-living amoebae. Pathogenic mycobacterial species can replicate in amoebae, whereas *M. smegmatis* cannot (Cirillo et al., 1997). Moreover, to the extent tested, the same mycobacterial

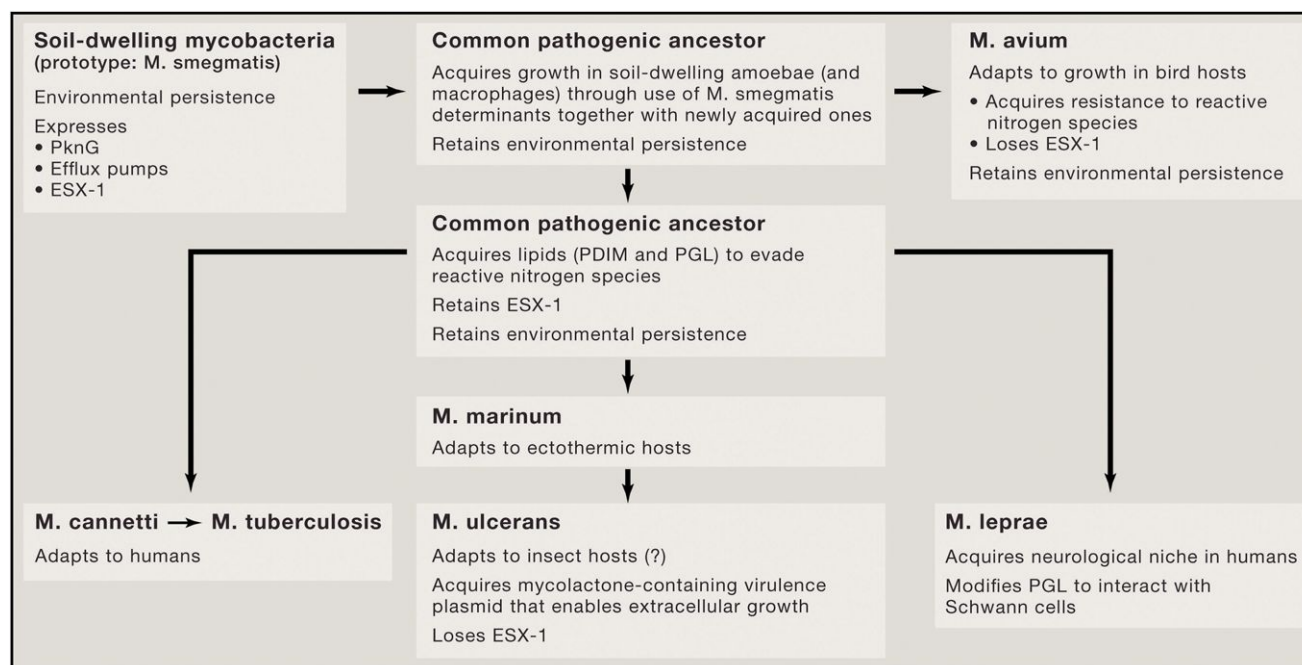


Figure 3. An Evolutionary Perspective of Mycobacterial Pathogenicity

An analysis of the pathogenic traits and preferred hosts of diverse mycobacterial species in relation to their genomes.

determinants are required for growth in macrophages and amoebae (Alibaud et al., 2011; Hagedorn et al., 2009; Hagedorn and Soldati, 2007; Solomon et al., 2003). Thus, predatory environmental amoebae may have served as the ancient evolutionary training ground for mycobacterial pathogens to survive in the macrophages of their multicellular hosts. This has been postulated for *Legionella*, an accidental human pathogen that can cause serious pneumonia after being aerosolized from potable water sources where it is thought to be sustained through replication in environmental amoebae (Fields et al., 2002).

Any foreign particulate that is phagocytosed by macrophages is destined to be processed through the endocytic pathways. Thus, intracellular pathogens have evolved diverse ingenious signature strategies to thwart, modulate, exploit, or avoid host endocytic pathways. Broadly speaking, these pathogens can resist lysosomal fusion to reside in non-acidified endosome-like compartments, survive (or even require) acidification so as to be able to reside in acidified lysosome-like compartments, or break out of the phagosome altogether to reside in the cytosol (Alix et al., 2011) (Figure 4). Most experimental studies on the virulence of intracellular mycobacteria have been conducted using either mouse or human cultured macrophages. For *M. tuberculosis*, observations in cultured macrophages have produced disparate results probably because the cell lines, culture conditions, and kinetics of infection differ considerably between different laboratories. *M. tuberculosis* is reportedly found localized to non-acidified early endosomes or found in acidified lysosomes with a small proportion of the bacteria eventually breaking out of the phagosome to reside in the cytosol (Cosma et al., 2003; van der Wel et al., 2007). Indeed, mycobacteria

have specific virulence determinants that promote, at least in cultured cells, both the avoidance of acidification as well as acid resistance, suggesting that, despite their best attempts, they might find themselves in acidified compartments (Rohde et al., 2007). In addition, the ability to break out into the cytosol is dependent on a specialized bacterial secretion system, ESX-1 (van der Wel et al., 2007), whose role we will elaborate upon in the context of the tuberculous granuloma in the following section.

The multiple subcellular compartments that *M. tuberculosis* can occupy within macrophages speak to the plethora of defenses with which they must contend even in the most permissive of macrophages. It is hardly surprising that diverse mycobacterial determinants are required for macrophage survival (Forrellad et al., 2013). What is surprising is that the obvious prediction that these determinants were acquired during mycobacterium's jump to becoming an amoeba dweller does not stand scrutiny. Although our search was not exhaustive, virtually all of the important *M. tuberculosis* virulence determinants that specifically promote intracellular growth are present in *M. smegmatis* (Forrellad et al., 2013)! What is more, in the cases tested, the *M. smegmatis* gene can substitute for its *M. tuberculosis* gene in mediating macrophage growth and virulence, suggesting that no or few further modifications were needed to confer this function (Houben et al., 2009). For instance, the eukaryotic-like serine-threonine protein kinase PknG is secreted into the phagosomal lumen and promotes macrophage growth by inhibiting lysosomal fusion and thereby acidification of the mycobacterial phagosome (Walburger et al., 2004). Although the *M. smegmatis* PknG homolog is able to restore macrophage growth of the *M. tuberculosis* *pknG*

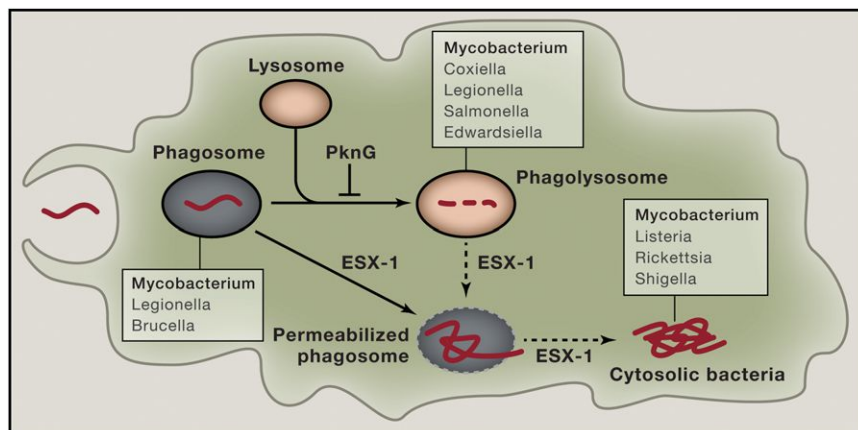


Figure 4. Intracellular Niches of *M. tuberculosis*

The observed intracellular niches of *M. tuberculosis* within macrophages are shown with other pathogens occupying those niches also listed. Confirmed trafficking pathways are indicated with continuous arrows and putative ones with dashed arrows. Pathways dependent on the mycobacterial ESX1 secretion system are indicated.

inhibitors should prevent the replication of intracellular *M. tuberculosis*. Moreover, combining efflux pump inhibitors with standard antibiotic therapy should be a “double whammy” to intracellular *M. tuberculosis* because they should

mutant, its function in the context of this saprophytic organism remains unknown. PknG is translationally repressed in *M. smegmatis* at least under axenic growth, suggesting that there is some specific “real-life” situation during life in the soil when it is induced, presumably to perform some specific function (Houben et al., 2009).

Perhaps the most fascinating example of conservation across mycobacteria is that of the mycobacterial energy-dependent efflux pumps, which we recently discovered to be *M. tuberculosis* macrophage growth factors by a circuitous route when looking for the basis of antibiotic tolerance (Adams et al., 2011; Rengarajan et al., 2005; Szumowski et al., 2013). In addition to developing genetic drug resistance through fixed mutations, *M. tuberculosis* famously develops what is called “phenotypic drug resistance” or “drug tolerance,” wherein it becomes transiently resistant to antibiotics (in the absence of fixed genetic mutations) in the host. This necessitates long treatment periods to achieve clinical cures (Connolly et al., 2007). Mycobacterial drug tolerance has long been attributed to the bacteria being in a non-replicating or dormant state in the host (Chao and Rubin, 2010; Rittershaus et al., 2013). However, our recent work shows that, when *M. tuberculosis* enters macrophages, it is, in fact, the actively replicating bacteria within the macrophages that develop antibiotic tolerance through the induction of specific macrophage-induced efflux pumps (Adams et al., 2011; Schnappinger et al., 2003). These same efflux pumps that mediate macrophage-induced drug tolerance also promote intracellular mycobacterial growth (Adams et al., 2011; Schnappinger et al., 2003). This suggests that these pumps may have evolved in the soil dwellers to defend against environmental toxins and inhibitors (including naturally occurring antibiotics) but came to be useful for the contemporary lifestyle of the pathogenic mycobacteria, facilitating intracellular survival, perhaps by protecting them against the antimicrobial peptides in macrophages. With the advent of chemotherapy, their ancestral function to defend against antibiotics or other growth inhibitors affords added benefit in surviving within the host; the same pumps may efflux the natural macrophage defenses as well as the administered antibiotics (Adams et al., 2011; Schnappinger et al., 2003). These findings additionally have therapeutic implications because efflux pump

inhibit intracellular growth in their own right and additionally allow the antibiotics to kill these bacteria better by preventing their efflux. Indeed, we have found this to be the case using inexpensive, well-tolerated approved human drugs that are currently used for other purposes (e.g., verapamil) (Adams et al., 2011, 2014). Moreover, the addition of verapamil to standard antituberculous chemotherapy reduces relapse rates in *M. tuberculosis*-infected mice (Gupta et al., 2013). On the basis of all these findings, clinical trials of verapamil as a TB treatment-shortening agent are imminent.

In summary, returning to our comparative theme, it would appear that the environmental mycobacteria, e.g., *M. smegmatis*, had determinants that allowed them to survive in the soil even though they could not survive within unicellular predators. The selection for the ability to survive within unicellular amoebae and eons later within the macrophages of multicellular creatures required using these determinants together with acquiring new as yet unknown ones, possibly by horizontal gene transfer (Figure 4). One could argue that mycobacteria come “pre-loaded” with the means to survive within a professional phagocytic cell. It is perhaps a general strategy for other soil organisms adapting to animal hosts as illustrated as well for *Rhodococcus equi*, the horse and occasional human pathogen that diversified from soil-dwelling *Rhodococci* (Letek et al., 2010). In the case of mycobacteria, the pathogenic forbearer selected for growth in amoeba appears to have then followed different evolutionary branches to dwell in different hosts, sometimes involving the acquisition of plasmids and, as is so often the case, gene loss, so as to fine-tune adaptation to specific hosts (Boritsch et al., 2014; Wang and Behr, 2014) (Figure 3).

In the following section, we will discuss the elaborate macrophage manipulation strategies used by *M. tuberculosis* to form the granuloma—the hallmark pathological structure of TB. We argue that it is as much a mycobacterial strategy for survival as it is a host defense response.

Multiplication in the Host—Exploiting the Granuloma

A granuloma is fundamentally an organized aggregate of macrophages whose membranes become tightly interdigitated like those of epithelial cells, leading them to be called epithelioid cells

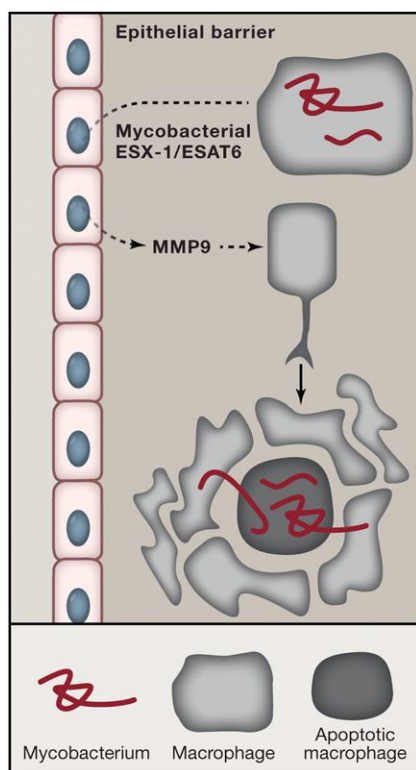


Figure 5. Mycobacteria Exploit the Granuloma to Expand Their Numbers in Early Infection

Mycobacteria within infected macrophages induce in an ESX1-dependent fashion MMP9 expression in epithelial cells surrounding the nascent granuloma. MMP9 stimulates the recruitment of new macrophages to the granuloma. Multiple new arrivals phagocytose the bacterial contents of a given dying infected macrophage, thus spreading the bacteria to new macrophages and providing them new expansion niches.

(Adams, 1976; Bouley et al., 2001). Granulomas can form in response to any number of persistent stimuli, both infectious and noninfectious, so that they are associated with myriad diseases (Ramakrishnan, 2012). They were first recognized as distinct structural entities in the context of human TB in the 17th century, preceding by some 200 years the discovery of *M. tuberculosis* as its cause, and even today, TB remains the most common cause of granulomas worldwide (Ramakrishnan, 2012). For a very long time, the tuberculous granuloma has been held to be an essential host protective structure—a fortress containing a complex mixture of diverse host cells that walls off bacteria (Saunders and Cooper, 2000) (Chao and Rubin, 2010; Rittershaus et al., 2013). Indeed, clinical and epidemiological studies clearly support the idea that the granuloma can sterilize infection in many cases (Cosma et al., 2004; Feldman and Baggenstoss, 1938). Yet, we argue that there is an inextricable link between highly organized granulomas and heavy bacterial burdens in TB, suggesting that, in many cases, the granuloma can be at least conducive to high bacterial burdens if not downright supportive of them (Connolly et al., 2007). Indeed, our findings over the last decade suggest that mycobacteria actually enhance the formation of granulomas and have adapted to

exploit these structures for their expansion and dissemination (Davis and Ramakrishnan, 2009). This modified view of the granuloma was made possible by studies in the zebrafish in which we could visualize the earliest events of granuloma formation around the first infected macrophage that had arrived in the deep tissues. Bacterial expansion in granulomas is accomplished by the spread of bacteria from dying macrophages to newly arriving ones (Davis and Ramakrishnan, 2009). When mycobacterial numbers increase to a certain threshold in individual macrophages, they undergo an apoptotic death that leaves viable bacteria still encased within the dead cells (Figure 5). Concomitantly, multiple new uninfected macrophages are recruited to the nascent granuloma that engulf the bacterial contents of a given dead or dying macrophage, thus enabling them to fill up the new cells. This process of macrophage death and re-phagocytosis enables a tremendous expansion of the bacterial niche, all within macrophages (Davis and Ramakrishnan, 2009).

At a molecular level, this coordinated macrophage death and re-phagocytosis is mediated through a specialized mycobacterial secretion system called ESX-1, most likely through its secreted effector ESAT6 (Davis and Ramakrishnan, 2009; Volkman et al., 2004; Volkman et al., 2010) (Figure 5). ESAT6 has been shown to induce apoptosis of infected cells in culture through multiple pathways, one or more of which may be operant in the granuloma (Choi et al., 2010; Derrick and Morris, 2007; Keane et al., 1997; Mishra et al., 2010; Swaim et al., 2006). ESX-1/ESAT-6 also recruits macrophages by inducing host matrix metalloproteinase 9 (MMP9) in epithelial cells surrounding the nascent granuloma (Volkman et al., 2010). If host MMP9 function is decreased, infection is attenuated, with reduced granuloma formation (Volkman et al., 2010). These discoveries, initially made in the zebrafish, are corroborated by findings in human TB showing that MMP9 is induced in epithelial cells surrounding lung granulomas, and increased MMP9 secretion is associated with increased severity and mortality in tuberculous meningitis (Elkington et al., 2007; Price et al., 2001).

In summary, *M. tuberculosis* appears to use at least two distinct pathways to recruit macrophages. When *M. tuberculosis* first enters the host animal, it uses its PGL surface lipid to recruit macrophages through host CCL2, which then bring the mycobacteria across host epithelium to deeper tissues (Cambier et al., 2014) (Figure 2). The intracellular bacteria then orchestrate the recruitment of additional macrophages to form the granuloma through their ESX-1 locus (Figure 5). Why the bacteria transition from using PGL to using ESAT6 to drive macrophage recruitment is unclear. What is clear is that, in both phases, macrophage recruitment benefits the mycobacteria as much as the host! Thus, CCL2 and MMP9 may both represent host determinants that have been co-opted by mycobacteria for their benefit, and increased CCL2 and MMP9 expression are both linked to human susceptibility to TB (Elkington et al., 2007; Price et al., 2001; Flores-Villanueva et al., 2005).

It is curious that, although the growing granuloma supports *M. tuberculosis* expansion, the macrophages within this structure generally become more microbicidal suggesting that the bacterium should be put at a further disadvantage (Adams, 1976; Bouley et al., 2001). But it appears that the mycobacteria,

in turn, rapidly adapt to the more hostile environment in the granuloma by transcriptionally inducing new genes (e.g., efflux pumps) upon entering a macrophage, which aid in intracellular survival. Moreover, when an infected macrophage forms or joins a granuloma, its intracellular bacteria rapidly induce additional genes that help it to counter the additional stresses of living within the cellular environment of the granuloma (Cosma et al., 2004; Davis et al., 2002; Ramakrishnan et al., 2000). If an infected animal with mature granulomas (frogs or zebrafish with *M. marinum*, or mice with *M. tuberculosis*) is superinfected with new mycobacteria, these bacteria enter new macrophages that preferentially migrate to existing granulomas rather than avoiding them as hostile sites (Cosma et al., 2004, 2008). It is as if mycobacteria “know” that, although the granuloma may not be a place where “the living is easy,” it still does afford them a preferred and, paradoxically, a protected multiplication niche and, as we shall see in the following section, a transmission niche as well.

Our proposed mycobacterial-centric view of granuloma biology and function returns us yet again to questions about how mycobacteria evolved to pathogenicity. We focus now on the ESX-1 secretion system and its effector, ESAT6. ESX-1/ESAT6, like most other *M. tuberculosis* virulence factors, is also present in *M. smegmatis*. However, in *M. smegmatis*, this secretion system regulates bacterial conjugation (Coros et al., 2008; Parsons et al., 1998) (Figure 3). The *M. tuberculosis* and *M. smegmatis* homologs are functionally conserved (Converse and Cox, 2005; Flint et al., 2004), but the co-option of bacterial gene-exchange systems for virulence is not special to mycobacteria. In diverse bacterial pathogens like *H. pylori*, *Legionella*, and *Agrobacterium*, the type IV secretion system that is required for virulence also mediates DNA transfer or uptake. Equally interesting is the finding that ESX is not even mycobacterium specific; an ESX homolog mediates virulence in *R. equi* while also being present in the soil *Rhodococci* and, indeed, being widely distributed in GC-rich soil bacteria (Letek et al., 2010). The mechanistic details of how ESX-1/ESAT6 mediates granuloma formation and virulence remain to be understood. ESAT6 is reported to be associated with the ability of *M. tuberculosis* subpopulations to break out of the phagosome (De Leon et al., 2012; Hsu et al., 2003; van der Wel et al., 2007) (Figure 4). In fact, prior to full-fledged phagosomal rupture, ESAT6 may permeabilize the phagosomal membrane enough to expose mycobacterial DNA to the host cytosolic DNA-sensing pathway (Manzanillo et al., 2012). This results in the induction of type I interferon, a cytokine that is best known for its antiviral activity; however, mycobacteria appear to drive this cytosolic DNA-sensing response in their favor. Components of the cytosolic pathway and type I interferon mediate host susceptibility in animal models. Type I interferon receptor knockout mice are more protected against TB, similar to the protection against TB afforded by knocking out MMP9 (Manzanillo et al., 2012; Mayer-Barber et al., 2014; Taylor et al., 2006; Watson et al., 2012). In humans too, a type I IFN transcriptional signature has been associated with active versus subclinical TB, suggesting it is a host susceptibility factor for disease progression (Berry et al., 2010). Whether ESAT6 modulates host MMP9 activity and granuloma formation through its membrane permeabilization activity and/or by triggering the cytosolic DNA

detection system still remains unclear. What is clear is that it represents another fascinating case in which a ubiquitous bacterial determinant has been co-opted and fine-tuned through genetic selection—not just to modulate bacterial growth in macrophages, but also to become integrated with other bacterial genes for a choreographed manipulation of macrophage migration and death to build the granuloma.

Of course, this initial phase of bacterial expansion in the granuloma is followed by the advent of an adaptive immune response that can often eradicate the tubercle bacilli, presumably by increasing the microbicidal capacity of the granuloma macrophages (Ramakrishnan, 2012). Epidemiological evidence would suggest that, with a little help from adaptive immunity, the granuloma can sterilize infection in most cases (Cosma et al., 2004; Feldman and Baggenstoss, 1938). Conversely, the very large number of active TB cases with full-fledged mature granulomas that are replete with lymphocytes suggests that mycobacteria have evolved additional strategies to evade adaptive immunity. As we have detailed in prior reviews, many of these strategies have been elucidated recently and include delaying both T cell priming in the lymph nodes and their arrival and activity in the granuloma (Pagan and Ramakrishnan, 2014; Ramakrishnan, 2012). Thus, *M. tuberculosis* reduces macrophage responsiveness to signaling by γ interferon, the main T cell cytokine (Banaiee et al., 2006). Finally, *M. tuberculosis* can synthesize its own tryptophan so that, unlike some other intracellular pathogens (e.g., Chlamydia), it is able to survive the intracellular tryptophan starvation brought on by γ interferon (Zhang et al., 2013). Thus, bacterial interactions with the host adaptive immune system add layers of complexity to the host-pathogen interface. The tubercle bacillus first induces an innate inflammatory response to accelerate macrophage responses (recruitment, phagocytosis, apoptosis) that are normally protective to turn the granuloma response into a bacterial production factory. By then delaying T cell priming, arrival, and activation together with macrophage responsiveness by what appears to be a highly orchestrated strategy, the bacteria buy themselves yet more time to establish a strong replicative niche in the granuloma.

Many diverse microorganisms induce granulomas, and it will be interesting to compare the pathways and consequences of granuloma formation for each of them. Within the mycobacteria, a tantalizing difference in granuloma formation pathways is already apparent by comparing *M. marinum* and *M. tuberculosis* to *M. avium* (Figure 3). ESAT6/ESX1 is absent from *M. avium*, which is not a particularly successful pathogen in human hosts (Dhama et al., 2011; Dumarey et al., 1994; Gomes et al., 1999; Onwueme et al., 2005; Sørensen et al., 1995) (Figure 3). However, in birds, *M. avium* causes a full-blown granulomatous disease that is transmissible, usually by ingestion. So *M. avium* granulomas must form through a different pathway. Meanwhile, *M. tuberculosis* ESAT6 has been recently revealed to have yet another function in supporting bacterial expansion in the granuloma—it induces immunosuppressive regulatory T cell populations that delay the migration of effector T cells into the granuloma (Shafiani et al., 2013; Shafiani et al., 2010). Thus, ESAT6 may prolong the phase of bacterial expansion in the innate granuloma that it orchestrates in the first place (Davis and Ramakrishnan, 2009). In this context, it is intriguing

that birds, in contrast to mammals and zebrafish, appear to have lost the transcription factor FoxP3 that is required for regulatory T cell development (Andersen et al., 2012). Therefore, a plausible explanation for *M. avium*'s loss of ESX-1/ESAT6 is that it is superfluous for granuloma expansion in its bird hosts. In contrast, its retention in *M. tuberculosis* and *M. marinum* may reflect its central role in granuloma expansion in mammalian and fish hosts. The selective forces of the host's immune system are reflected in the array of expressed virulence genes seen in each host-adapted mycobacterial species. Although our conjecture on the evolutionary connections may well be incorrect, our revised view of the role of the granuloma in the pathogenesis of tuberculosis can be tested further experimentally and may have potential clinical relevance. We predict that the reduction of granuloma formation by pharmacological inhibition of the relevant host pathways (e.g., MMP9) should ameliorate infection.

Exit from the Host—Escaping the Granuloma

A host-adapted pathogen's final step to ensure its evolutionary success is to exit the host and enter a new host for the infection cycle to start anew. In the case of *M. tuberculosis*, it must exit the granuloma of its infected host to enter and establish infection in a new host. Epidemiological evidence suggests that transmission occurs most efficiently from individuals with organized granulomas that have undergone central necrosis (Sharma et al., 2005; Reichler et al., 2002; Bekker and Wood, 2010; Huang et al., 2014). The necrotic areas rupture into the bronchial tree, thus exposing the mycobacteria to the airway whence they can be aerosolized in cough droplets. Recent work from both our laboratory and others takes a slightly modified view of the dynamics of cell death in a granuloma (Ramakrishnan, 2012). Broadly speaking, infected granuloma macrophages can die in two ways: by apoptosis or by necrosis. Apoptosis leaves the host cell membranes intact so that the bacteria remain encased within the macrophage corpse and are readily phagocytosed by new entering cells. In contrast, macrophage necrosis, or lysis, releases the intact bacteria into the extracellular milieu. This necrotic debris, or caseum, seems to be an ideal bacterial growth medium as the multiplying bacteria reach much higher numbers extracellularly and grow in characteristic serpentine cords. Thus, in our view, apoptotic death favors bacterial expansion or maintenance of the granuloma by providing new cells to grow in, albeit still restricted by macrophage defenses. Bacterial lysis from the macrophages allows more exuberant growth, reflecting the ineffectiveness of extracellular host defenses against mycobacteria. Moreover, recent work suggests that these corded extracellular mycobacteria are not readily engulfed by new macrophages the way that bacteria within apoptotic cells are (Bernut et al., 2014). It is not fully understood how the dynamics of the granuloma shift to favor necrosis, but some interesting insights are emerging.

Our studies in *M. marinum*-infected zebrafish have uncovered two pathways that lead to macrophage necrosis, each through opposite dysregulation of TNF, resulting in too little or too much TNF. TNF is required for macrophage microbicidal activity, although we do not understand the specific mechanisms and effectors (Clay et al., 2008). Whereas host TNF deficiency causes mycobacteria to grow exuberantly within macrophages that then

die and release them to the extracellular milieu (Clay et al., 2008), we have found that an excess of host TNF causes infected macrophages to undergo necrosis, through a programmed pathway called necroptosis (Roca and Ramakrishnan, 2013; Tobin et al., 2012). Host TNF excess causes the activation of the RIP1 and RIP3 kinases that then, through a series of steps, induce reactive oxygen species in the macrophage mitochondria (Roca and Ramakrishnan, 2013; Tobin et al., 2012). Reactive oxygen has dual effects: it kills both mycobacteria and macrophage. The net result is that, just as the macrophage is on its way to killing its infecting mycobacteria, it dies. The few surviving mycobacteria that are released extracellularly can expand their numbers rapidly.

Perturbation of several pathways could lead to TNF dysregulation and, in turn, granuloma necrosis. One that we identified in a zebrafish mutant screen for susceptibility to *M. marinum* involves dysregulation of the leukotriene A4 hydrolase (LTA4H), a synthetic enzyme in the eicosanoid pathway that catalyzes the synthesis of the highly proinflammatory lipid leukotriene B₄ (Tobin et al., 2010, 2012). LTA4H deficiency prevents the synthesis of this leukotriene and instead causes the accumulation of the anti-inflammatory lipid lipoxin A₄, which represses the TNF response. In contrast, LTA4H excess causes an overproduction of leukotriene B₄ and an excess of TNF. In humans, a common *LTA4H* promoter variant regulates gene expression and homozygotes for both the low- and high-expression variants that are associated with low and high inflammation, respectively. Individuals with low- and high-expression variants of LTA4H get severe tuberculous meningitis, with a high mortality. In contrast, the heterozygotes, with an intermediate (presumably optimal) level of LTA4H expression, are protected. These findings, in turn, have therapeutic implications—in a Vietnamese tuberculous meningitis cohort, adjunctive treatment with the broadly immunosuppressive glucocorticoids, which are now routinely administered along with antitubercular chemotherapy, only prevented mortality of the high LTA4H group while possibly increasing mortality of the low LTA4H individuals (Tobin et al., 2012; Tobin et al., 2010). These results suggest that patient genotype-directed glucocorticoid treatment may optimize TB treatment.

Our detailed understanding of the two pathways through which excess TNF-induced mitochondrial reactive oxygen causes macrophage necrosis also suggests new approaches to TB therapy (Roca and Ramakrishnan, 2013; Tobin et al., 2012). Reactive oxygen causes the translocation of the matrix mitochondrial protein cyclophilin D to participate in the formation of a pore on the mitochondrial membrane, thus causing leakage of mitochondrial contents. Additionally, the reactive oxygen also causes overproduction of a cellular lipid called ceramide that induces necrosis through mechanisms that are not yet clear. We have identified currently available oral drugs that can block each of these pathways. Alisporivir, a drug in phase 3 clinical trials for another disease, blocks cyclophilin D, and desipramine, a tricyclic antidepressant, inhibits ceramide production. In the zebrafish, the combined use of these drugs allows the reactive oxygen to kill the bacteria without killing the macrophages and thereby converts the hypersusceptible state of TNF excess to hyperresistant. It is possible that these drugs will have a similar effect in humans who induce excess TNF during infection.

Necrosis from excessive TNF may be further exacerbated by the participation of adaptive immunity, as much of the TNF induced in TB is produced by T cells (Roach et al., 2002; Saunders et al., 2004). Perhaps *M. tuberculosis* has co-opted T cells into producing the TNF that causes macrophage necrosis and thus encourages transmission. Indeed, T cell epitopes of all known mycobacterial T cell antigens are reported to be hyper-conserved across strains globally, even more so than those of essential genes (Comas et al., 2010). This finding suggests that T cell recognition favors the survival and transmission of mycobacteria, arguably by inducing TNF-mediated necrosis as described above. The hyperconservation of T cell epitope regions of expressed mycobacterial antigens highlights the close evolutionary relationship between T cell recognition and bacterial fitness. Epidemiological evidence suggests that individuals with diminished adaptive immunity (e.g., HIV-infected individuals or children) tend to have smaller and less necrotic granulomas than immunocompetent adults, and these former individuals—though more susceptible to TB—do not transmit it as well as the latter (Huang et al., 2014). Globally, >80% of active TB occurs in individuals who are not HIV infected (Zumla et al., 2013). This link between disease and an intact host adaptive immune system may suggest that *M. tuberculosis* takes advantage of adaptive immunity for its transmission.

In summary, our view is that *M. tuberculosis* and many other pathogenic mycobacteria are not innocent bystanders during the formation of granuloma. They can modulate the host response to infection to build and modify this complex immunological entity into a niche that can sustain infection, first through intracellular growth and then through extracellular growth that also favors transmission.

Concluding Thoughts

The case of *M. tuberculosis* exemplifies how a series of genetic adaptations can convert a soil-dwelling microbe into one of the most successful and enduring pathogens of humanity. More people are thought to have died of TB than of any other infectious disease throughout history, and more people are afflicted with active TB disease today than at any other time in history (Lawn and Zumla, 2011). While marveling at the exquisite bacterial adaptations that have honed this microbe's success in its human niche, it is important to remember that most infected individuals (classically reported to be 90%) can successfully contain or clear the infection (Zumla et al., 2013). This occurs either through an initial mobilization of innate immune mechanisms or, failing that, through adaptive immunity. In this Review, we have tried to point out how the outcome of each step of the host-pathogen interaction can represent “success” for the host—infection can be suppressed or cleared at the first site of infection, in the innate granuloma, or later, when the granuloma is further re-enforced by adaptive immunity (Cambier et al., 2014; Lin et al., 2014; Adams et al., 2011; Rengarajan et al., 2005; Szumowski et al., 2013).

Suppression of infection can result in a clinical latency during which the bacteria persist indefinitely in the host and can produce active disease even decades later—a scenario that is emphasized in the literature (Chao and Rubin, 2010; Cosma et al., 2003; Rittershaus et al., 2013). However, careful longitudinal studies from the pre-antibiotic era suggest that most

contemporary human disease manifests within a few months of infection or is cleared (Cosma et al., 2003). In today's world, it is these relatively recently infected individuals who transmit the bulk of the disease, rather than those in whom disease has recrudesced after many decades. From a medical perspective, this places the impetus on understanding how the majority of infected individuals progress to disease relatively rapidly. However, it is likely that clinical latency played an important role in sustaining the organism through the ~60,000 years before TB became a “crowd” disease. The organism early on in its evolution did not have the advantage of large susceptible populations that resulted from the Neolithic revolution of domestication and the development of agriculture. It is hard to imagine how the early hunter-gathers living in small groups could have sustained *M. tuberculosis* without the benefit of activation of transmissible disease in previously healthy infected individuals who were able to travel long distances. We speculate that, in the latter part of its history, TB has shifted from being a heritage disease to a crowd disease, and the opportunities afforded by a growing susceptible host population may have led not only to increased transmission, but also to a more aggressive stance against innate immune defenses, leading to epidemic spread rather than persistence.

It is interesting in this context that we may be witnessing a shift in transmission of another major mycobacterial disease leprosy. Leprosy is caused by *Mycobacterium leprae*, which appears to have also evolved from the common *M. tuberculosis*-*M. marinum* ancestor (Figure 3). *M. leprae* is particularly intriguing because it has undergone substantial gene reduction to the point where it has lost its capacity for axenic growth (Cole et al., 2001). At the same time, it has become specialized in its pathogenic niche, infecting Schwann cells of the peripheral nervous system through complex mechanisms (Masaki et al., 2013). For most of its pathogenic human history, *M. leprae* has been a strict human pathogen, with transmission occurring only through prolonged contact with infected humans. Yet, in very recent times, the nine banded armadillo in the Southeastern United States became infected from humans and has now become a full-fledged reservoir for disease, so that leprosy is now mainly a zoonotic disease in this area (Truman et al., 2011). Albeit in a less dramatic way, *M. tuberculosis* must clearly have made adaptations to the very great changes in human lifestyles to retain its success. Understanding these changes may have more than academic value; it may help us better understand the disease itself and hence its treatment. The parallel evolution of pathogens to keep up with changing environments is hardly unique to mycobacteria but is shared with many, many other pathogens.

Finally, our recent work has repeatedly confronted us with the fact that TB is not as much a disease of failed immunity as it is of coevolution. At every step of infection, the bacterium appears to be inducing and benefitting from an over-exuberant response using the very inflammatory pathways that are thought to have evolved to thwart bacteria—CCL2 induction to enter the host, MMP9 induction to expand in the granuloma, and finally TNF and T lymphocytes to exit the granuloma for transmission. Despite all of this, both host and pathogen have prevailed. That is why we consider *M. tuberculosis* to be a paradigm of a host-adapted microorganism. It has coevolved with the human immune system, discarding and gaining genes to be in tune

with it. As an ancient disease agent adapting to humans, the microbe could not have anticipated that humans would be so successful. Yet its stealth and subtlety have allowed it to thrive even in the face of modern medicine. While new therapeutic avenues will most likely be found, we must not undervalue the power of genetic selection for the survival of any microorganism—and, based on past performance, particularly for *M. tuberculosis*.

ACKNOWLEDGMENTS

We thank Kevin Takaki for help with designing and making the figures and Marcel Behr, Pamela Small, and David Tobin for their critique of the paper. C.J.C. was funded by an NIH bacterial pathogenesis training grant award. L.R. is a recipient of the NIH Director's Pioneer Award and a Wellcome Trust Principal Research Fellowship.

REFERENCES

- Adams, D.O. (1976). The granulomatous inflammatory response. A review. *Am. J. Pathol.* 84, 164–192.
- Adams, K.N., Takaki, K., Connolly, L.E., Wiedenhoft, H., Winglee, K., Humbert, O., Edelstein, P.H., Cosma, C.L., and Ramakrishnan, L. (2011). Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. *Cell* 145, 39–53.
- Adams, K.N., Szumowski, J.D., and Ramakrishnan, L. (2014). Verapamil, and its metabolite norverapamil, inhibit macrophage-induced, bacterial efflux pump-mediated tolerance to multiple anti-tubercular drugs. *J. Infect. Dis.* 210, 456–466.
- Alibaud, L., Rombouts, Y., Trivelli, X., Burguière, A., Cirillo, S.L., Cirillo, J.D., Dubremetz, J.F., Guérardel, Y., Lutfalla, G., and Kremer, L. (2011). A *Mycobacterium marinum* TesA mutant defective for major cell wall-associated lipids is highly attenuated in *Dictyostelium discoideum* and zebrafish embryos. *Mol. Microbiol.* 80, 919–934.
- Alix, E., Mukherjee, S., and Roy, C.R. (2011). Subversion of membrane transport pathways by vacuolar pathogens. *J. Cell Biol.* 195, 943–952.
- Andersen, K.G., Nissen, J.K., and Betz, A.G. (2012). Comparative Genomics Reveals Key Gain-of-Function Events in *Foxp3* during Regulatory T Cell Evolution. *Front. Immunol.* 3, 113.
- Banaiee, N., Kincaid, E.Z., Buchwald, U., Jacobs, W.R., Jr., and Ernst, J.D. (2006). Potent inhibition of macrophage responses to IFN- γ by live virulent *Mycobacterium tuberculosis* is independent of mature mycobacterial lipoproteins but dependent on TLR2. *J. Immunol.* 176, 3019–3027.
- Bates, J.H., Potts, W.E., and Lewis, M. (1965). Epidemiology of primary tuberculosis in an industrial school. *N. Engl. J. Med.* 272, 714–717.
- Bekker, L.G., and Wood, R. (2010). The changing natural history of tuberculosis and HIV coinfection in an urban area of hyperendemicity. *Clin. Infect. Dis.* 50 (Suppl 3), S208–S214.
- Belkaid, Y., and Hand, T.W. (2014). Role of the microbiota in immunity and inflammation. *Cell* 157, 121–141.
- Bernut, A., Herrmann, J.L., Kissa, K., Dubremetz, J.F., Gaillard, J.L., Lutfalla, G., and Kremer, L. (2014). *Mycobacterium abscessus* cording prevents phagocytosis and promotes abscess formation. *Proc. Natl. Acad. Sci. USA* 111, E943–E952.
- Berry, M.P.R., Graham, C.M., McNab, F.W., Xu, Z., Bloch, S.A.A., Oni, T., Wilkinson, K.A., Banchereau, R., Skinner, J., Wilkinson, R.J., et al. (2010). An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 466, 973–977.
- Boritsch, E.C., Supply, P., Honoré, N., Seeman, T., Stinear, T.P., and Brosch, R. (2014). A glimpse into the past and predictions for the future: the molecular evolution of the tuberculosis agent. *Mol. Microbiol.* 93, 835–852.
- Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., et al. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514, 494–497.
- Bouley, D.M., Ghorri, N., Mercer, K.L., Falkow, S., and Ramakrishnan, L. (2001). Dynamic nature of host-pathogen interactions in *Mycobacterium marinum* granulomas. *Infect. Immun.* 69, 7820–7831.
- Cambier, C.J., Takaki, K.K., Larson, R.P., Hernandez, R.E., Tobin, D.M., Urdahl, K.B., Cosma, C.L., and Ramakrishnan, L. (2014). *Mycobacteria* manipulate macrophage recruitment through coordinated use of membrane lipids. *Nature* 505, 218–222.
- Chao, M.C., and Rubin, E.J. (2010). Letting sleeping dogs lie: does dormancy play a role in tuberculosis? *Annu. Rev. Microbiol.* 64, 293–311.
- Charlson, E.S., Bittinger, K., Haas, A.R., Fitzgerald, A.S., Frank, I., Yadav, A., Bushman, F.D., and Collman, R.G. (2011). Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am. J. Respir. Crit. Care Med.* 184, 957–963.
- Choi, H.H., Shin, D.M., Kang, G., Kim, K.H., Park, J.B., Hur, G.M., Lee, H.M., Lim, Y.J., Park, J.K., Jo, E.K., and Song, C.H. (2010). Endoplasmic reticulum stress response is involved in *Mycobacterium tuberculosis* protein ESAT-6-mediated apoptosis. *FEBS Lett.* 584, 2445–2454.
- Cirillo, J.D., Falkow, S., Tompkins, L.S., and Bermudez, L.E. (1997). Interaction of *Mycobacterium avium* with environmental amoebae enhances virulence. *Infect. Immun.* 65, 3759–3767.
- Clay, H., Volkman, H.E., and Ramakrishnan, L. (2008). Tumor necrosis factor signaling mediates resistance to mycobacteria by inhibiting bacterial growth and macrophage death. *Immunity* 29, 283–294.
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, et al. (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544.
- Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honoré, N., Garnier, T., Churcher, C., Harris, D., et al. (2001). Massive gene decay in the leprosy bacillus. *Nature* 409, 1007–1011.
- Comas, I., Chakravarti, J., Small, P.M., Galagan, J., Niemann, S., Kremer, K., Ernst, J.D., and Gagneux, S. (2010). Human T cell epitopes of *Mycobacterium tuberculosis* are evolutionarily hyperconserved. *Nat. Genet.* 42, 498–503.
- Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., et al. (2013). Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat. Genet.* 45, 1176–1182.
- Connolly, L.E., Edelstein, P.H., and Ramakrishnan, L. (2007). Why is long-term therapy required to cure tuberculosis? *PLoS Med.* 4, e120.
- Converse, S.E., and Cox, J.S. (2005). A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*. *J. Bacteriol.* 187, 1238–1245.
- Coros, A., Callahan, B., Battaglioli, E., and Derbyshire, K.M. (2008). The specialized secretory apparatus ESX-1 is essential for DNA transfer in *Mycobacterium smegmatis*. *Mol. Microbiol.* 69, 794–808.
- Cosma, C.L., Sherman, D.R., and Ramakrishnan, L. (2003). The secret lives of the pathogenic mycobacteria. *Annu. Rev. Microbiol.* 57, 641–676.
- Cosma, C.L., Humbert, O., and Ramakrishnan, L. (2004). Superinfecting mycobacteria home to established tuberculous granulomas. *Nat. Immunol.* 5, 828–835.
- Cosma, C.L., Humbert, O., Sherman, D.R., and Ramakrishnan, L. (2008). Trafficking of superinfecting *Mycobacterium* organisms into established granulomas occurs in mammals and is independent of the *Erp* and *ESX-1* mycobacterial virulence loci. *J. Infect. Dis.* 198, 1851–1855.
- Davis, J.M., and Ramakrishnan, L. (2009). The role of the granuloma in expansion and dissemination of early tuberculous infection. *Cell* 136, 37–49.
- Davis, J.M., Clay, H., Lewis, J.L., Ghorri, N., Herbolme, P., and Ramakrishnan, L. (2002). Real-time visualization of mycobacterium-macrophage interactions leading to initiation of granuloma formation in zebrafish embryos. *Immunity* 17, 693–702.

- De Leon, J., Jiang, G., Ma, Y., Rubin, E., Fortune, S., and Sun, J. (2012). Mycobacterium tuberculosis ESAT-6 exhibits a unique membrane-interacting activity that is not found in its ortholog from non-pathogenic Mycobacterium smegmatis. *J. Biol. Chem.* 287, 44184–44191.
- Derrick, S.C., and Morris, S.L. (2007). The ESAT6 protein of Mycobacterium tuberculosis induces apoptosis of macrophages by activating caspase expression. *Cell. Microbiol.* 9, 1547–1555.
- Dhama, K., Mahendran, M., Tiwari, R., Dayal Singh, S., Kumar, D., Singh, S., and Sawant, P.M. (2011). Tuberculosis in Birds: Insights into the Mycobacterium avium Infections. *Vet. Med. Int.* 2011, 712369.
- Dumarey, C.H., Labrousse, V., Rastogi, N., Vargaftig, B.B., and Bachelet, M. (1994). Selective Mycobacterium avium-induced production of nitric oxide by human monocyte-derived macrophages. *J. Leukoc. Biol.* 56, 36–40.
- Elkington, P.T., Green, J.A., Emerson, J.E., Lopez-Pascua, L.D., Boyle, J.J., O'Kane, C.M., and Friedland, J.S. (2007). Synergistic up-regulation of epithelial cell matrix metalloproteinase-9 secretion in tuberculosis. *Am. J. Respir. Cell Mol. Biol.* 37, 431–437.
- Falkow, S. (2006). Is persistent bacterial infection good for your health? *Cell* 124, 699–702.
- Falkow, S. (2008). I never met a microbe I didn't like. *Nat. Med.* 14, 1053–1057.
- Fang, F.C. (2004). Antimicrobial reactive oxygen and nitrogen species: concepts and controversies. *Nat. Rev. Microbiol.* 2, 820–832.
- Feldman, W.H., and Baggenstoss, A.H. (1938). The residual infectivity of the primary complex of tuberculosis. *Am. J. Pathol.* 14, 473–490.
- Fields, B.S., Benson, R.F., and Besser, R.E. (2002). Legionella and Legionnaires' disease: 25 years of investigation. *Clin. Microbiol. Rev.* 15, 506–526.
- Flint, J.L., Kowalski, J.C., Karnati, P.K., and Derbyshire, K.M. (2004). The RD1 virulence locus of Mycobacterium tuberculosis regulates DNA transfer in Mycobacterium smegmatis. *Proc. Natl. Acad. Sci. USA* 101, 12598–12603.
- Flores-Villanueva, P.O., Ruiz-Morales, J.A., Song, C.-H., Flores, L.M., Jo, E.-K., Montañó, M., Barnes, P.F., Selman, M., and Granados, J. (2005). A functional promoter polymorphism in monocyte chemoattractant protein-1 is associated with increased susceptibility to pulmonary tuberculosis. *J. Exp. Med.* 202, 1649–1658.
- Forrellad, M.A., Klepp, L.I., Gioffré, A., Sabio y García, J., Morbidoni, H.R., de la Paz Santangelo, M., Cataldi, A.A., and Bigi, F. (2013). Virulence factors of the Mycobacterium tuberculosis complex. *Virulence* 4, 3–66.
- Gewirtz, A.T., Yu, Y., Krishna, U.S., Israel, D.A., Lyons, S.L., and Peek, R.M., Jr. (2004). Helicobacter pylori flagellin evades toll-like receptor 5-mediated innate immunity. *J. Infect. Dis.* 189, 1914–1920.
- Glickman, M.S., and Jacobs, W.R., Jr. (2001). Microbial pathogenesis of Mycobacterium tuberculosis: dawn of a discipline. *Cell* 104, 477–485.
- Gomes, M.S., Flórido, M., Pais, T.F., and Appelberg, R. (1999). Improved clearance of Mycobacterium avium upon disruption of the inducible nitric oxide synthase gene. *J. Immunol.* 162, 6734–6739.
- Gupta, S., Tyagi, S., Almeida, D.V., Maiga, M.C., Ammerman, N.C., and Bishai, W.R. (2013). Acceleration of tuberculosis treatment by adjunctive therapy with verapamil as an efflux inhibitor. *Am. J. Respir. Crit. Care Med.* 188, 600–607.
- Hagedorn, M., and Soldati, T. (2007). Flotillin and RacH modulate the intracellular immunity of Dictyostelium to Mycobacterium marinum infection. *Cell. Microbiol.* 9, 2716–2733.
- Hagedorn, M., Rohde, K.H., Russell, D.G., and Soldati, T. (2009). Infection by tubercular mycobacteria is spread by nonlytic ejection from their amoeba hosts. *Science* 323, 1729–1733.
- Henard, C.A., and Vázquez-Torres, A. (2011). Nitric oxide and salmonella pathogenesis. *Front. Microbiol.* 2, 84.
- Houben, E.N., Walburger, A., Ferrari, G., Nguyen, L., Thompson, C.J., Miess, C., Vogel, G., Mueller, B., and Pieters, J. (2009). Differential expression of a virulence factor in pathogenic and non-pathogenic mycobacteria. *Mol. Microbiol.* 72, 41–52.
- Houk, V.N. (1980). Spread of tuberculosis via recirculated air in a naval vessel: the Byrd study. *Ann. N Y Acad. Sci.* 353, 10–24.
- Hsu, T., Hingley-Wilson, S.M., Chen, B., Chen, M., Dai, A.Z., Morin, P.M., Marks, C.B., Padiyar, J., Goulding, C., Gingery, M., et al. (2003). The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. *Proc. Natl. Acad. Sci. USA* 100, 12420–12425.
- Huang, C.C., Tchetgen, E.T., Becerra, M.C., Cohen, T., Hughes, K.C., Zhang, Z., Calderon, R., Yataco, R., Contreras, C., Galea, J., et al. (2014). The effect of HIV-related immunosuppression on the risk of tuberculosis transmission to household contacts. *Clin. Infect. Dis.* 58, 765–774.
- Jones, B.D., Ghori, N., and Falkow, S. (1994). Salmonella typhimurium initiates murine infection by penetrating and destroying the specialized epithelial M cells of the Peyer's patches. *J. Exp. Med.* 180, 15–23.
- Keane, J., Balcewicz-Sablinska, M.K., Remold, H.G., Chupp, G.L., Meek, B.B., Fenton, M.J., and Kornfeld, H. (1997). Infection by Mycobacterium tuberculosis promotes human alveolar macrophage apoptosis. *Infect. Immun.* 65, 298–304.
- Kirksey, M.A., Tischler, A.D., Siméone, R., Hisert, K.B., Uplekar, S., Guilhot, C., and McKinney, J.D. (2011). Spontaneous phthiocerol dimycocerosate-deficient variants of Mycobacterium tuberculosis are susceptible to gamma interferon-mediated immunity. *Infect. Immun.* 79, 2829–2838.
- Lawn, S.D., and Zumla, A.I. (2011). Tuberculosis. *Lancet* 378, 57–72.
- Letek, M., González, P., Macarthur, I., Rodríguez, H., Freeman, T.C., Valero-Rello, A., Blanco, M., Buckley, T., Cherevach, I., Fahey, R., et al. (2010). The genome of a pathogenic rhodococcus: cooptive virulence underpinned by key gene acquisitions. *PLoS Genet.* 6, e1001145.
- Lin, P.L., Ford, C.B., Coleman, M.T., Myers, A.J., Gawande, R., Ioerger, T., Sacchettini, J., Fortune, S.M., and Flynn, J.L. (2014). Sterilization of granulomas is common in active and latent tuberculosis despite within-host variability in bacterial killing. *Nat. Med.* 20, 75–79.
- Liu, P.T., Stenger, S., Li, H., Wenzel, L., Tan, B.H., Krutzik, S.R., Ochoa, M.T., Schaub, J., Wu, K., Meinken, C., et al. (2006). Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. *Science* 311, 1770–1773.
- Manzanillo, P.S., Shiloh, M.U., Portnoy, D.A., and Cox, J.S. (2012). Mycobacterium tuberculosis activates the DNA-dependent cytosolic surveillance pathway within macrophages. *Cell Host Microbe* 11, 469–480.
- Masaki, T., Qu, J., Cholewa-Waclaw, J., Burr, K., Raaum, R., and Rambukana, A. (2013). Reprogramming adult Schwann cells to stem cell-like cells by leprosy bacilli promotes dissemination of infection. *Cell* 152, 51–67.
- Mayer-Barber, K.D., Andrade, B.B., Oland, S.D., Amaral, E.P., Barber, D.L., Gonzales, J., Derrick, S.C., Shi, R., Kumar, N.P., Wei, W., et al. (2014). Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. *Nature* 511, 99–103.
- Mishra, B.B., Moura-Alves, P., Sonawane, A., Hacohen, N., Griffiths, G., Moita, L.F., and Anes, E. (2010). Mycobacterium tuberculosis protein ESAT-6 is a potent activator of the NLRP3/ASC inflammasome. *Cell. Microbiol.* 12, 1046–1063.
- Monack, D.M. (2013). Helicobacter and salmonella persistent infection strategies. *Cold Spring Harb. Perspect. Med.* 3, a010348.
- Moran, A.P. (2007). Lipopolysaccharide in bacterial chronic infection: insights from Helicobacter pylori lipopolysaccharide and lipid A. *Int. J. Med. Microbiol.* 297, 307–319.
- Onwueme, K.C., Vos, C.J., Zurita, J., Ferreras, J.A., and Quadri, L.E.N. (2005). The dimycocerosate ester polyketide virulence factors of mycobacteria. *Prog. Lipid Res.* 44, 259–302.
- Pagan, A., and Ramakrishnan, L. (2014). Immunity and immunopathology in the the tuberculous granuloma. *Cold Spring Harb. Perspect. Med.* Published online November 6, 2014. <http://dx.doi.org/10.1101/cshperspect.a018499>.
- Parsons, L.M., Jankowski, C.S., and Derbyshire, K.M. (1998). Conjugal transfer of chromosomal DNA in Mycobacterium smegmatis. *Mol. Microbiol.* 28, 571–582.
- Price, N.M., Farrar, J., Tran, T.T., Nguyen, T.H., Tran, T.H., and Friedland, J.S. (2001). Identification of a matrix-degrading phenotype in human tuberculosis in vitro and in vivo. *J. Immunol.* 166, 4223–4230.
- Ramakrishnan, L. (2012). Revisiting the role of the granuloma in tuberculosis. *Nat. Rev. Immunol.* 12, 352–366.

- Ramakrishnan, L. (2013). The zebrafish guide to tuberculosis immunity and treatment. *Cold Spring Harb. Symp. Quant. Biol.* 78, 179–192.
- Ramakrishnan, L., Federspiel, N.A., and Falkow, S. (2000). Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family. *Science* 288, 1436–1439.
- Reed, M.B., Domenech, P., Manca, C., Su, H., Barczak, A.K., Kreiswirth, B.N., Kaplan, G., and Barry, C.E., 3rd. (2004). A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* 431, 84–87.
- Reichler, M.R., Reves, R., Bur, S., Thompson, V., Mangura, B.T., Ford, J., Valway, S.E., and Onorato, I.M.; Contact Investigation Study Group (2002). Evaluation of investigations conducted to detect and prevent transmission of tuberculosis. *JAMA* 287, 991–995.
- Rengarajan, J., Bloom, B.R., and Rubin, E.J. (2005). Genome-wide requirements for *Mycobacterium tuberculosis* adaptation and survival in macrophages. *Proc. Natl. Acad. Sci. USA* 102, 8327–8332.
- Rittershaus, E.S., Baek, S.H., and Sasseti, C.M. (2013). The normalcy of dormancy: common themes in microbial quiescence. *Cell Host Microbe* 13, 643–651.
- Roach, D.R., Bean, A.G.D., Demangel, C., France, M.P., Briscoe, H., and Britton, W.J. (2002). TNF regulates chemokine induction essential for cell recruitment, granuloma formation, and clearance of mycobacterial infection. *J. Immunol.* 168, 4620–4627.
- Roca, F.J., and Ramakrishnan, L. (2013). TNF dually mediates resistance and susceptibility to mycobacteria via mitochondrial reactive oxygen species. *Cell* 153, 521–534.
- Rohde, K., Yates, R.M., Purdy, G.E., and Russell, D.G. (2007). *Mycobacterium tuberculosis* and the environment within the phagosome. *Immunol. Rev.* 219, 37–54.
- Saunders, B.M., and Cooper, A.M. (2000). Restraining mycobacteria: role of granulomas in mycobacterial infections. *Immunol. Cell Biol.* 78, 334–341.
- Saunders, B.M., Briscoe, H., and Britton, W.J. (2004). T cell-derived tumour necrosis factor is essential, but not sufficient, for protection against *Mycobacterium tuberculosis* infection. *Clin. Exp. Immunol.* 137, 279–287.
- Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C., and Schoolnik, G.K. (2003). Transcriptional Adaptation of *Mycobacterium tuberculosis* within Macrophages: Insights into the Phagosomal Environment. *J. Exp. Med.* 198, 693–704.
- Shafiani, S., Tucker-Heard, G., Kariyone, A., Takatsu, K., and Urdahl, K.B. (2010). Pathogen-specific regulatory T cells delay the arrival of effector T cells in the lung during early tuberculosis. *J. Exp. Med.* 207, 1409–1420.
- Shafiani, S., Dinh, C., Ertelt, J.M., Moguche, A.O., Siddiqui, I., Smigiel, K.S., Sharma, P., Campbell, D.J., Way, S.S., and Urdahl, K.B. (2013). Pathogen-specific Treg cells expand early during mycobacterium tuberculosis infection but are later eliminated in response to Interleukin-12. *Immunity* 38, 1261–1270.
- Sharma, S.K., Mohan, A., Sharma, A., and Mitra, D.K. (2005). Miliary tuberculosis: new insights into an old disease. *Lancet Infect. Dis.* 5, 415–430.
- Shepard, C.C. (1957). Growth characteristics of tubercle bacilli and certain other mycobacteria in HeLa cells. *J. Exp. Med.* 105, 39–48.
- Siegrist, M.S., and Bertozzi, C.R. (2014). Mycobacterial lipid logic. *Cell Host Microbe* 15, 1–2.
- Solomon, J.M., Leung, G.S., and Isberg, R.R. (2003). Intracellular replication of *Mycobacterium marinum* within *Dictyostelium discoideum*: efficient replication in the absence of host coronin. *Infect. Immun.* 71, 3578–3586.
- Sorensen, A.L., Nagai, S., Houen, G., Andersen, P., and Andersen, A.B. (1995). Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. *Infect. Immun.* 63, 1710–1717.
- Stinear, T.P., Seemann, T., Harrison, P.F., Jenkin, G.A., Davies, J.K., Johnson, P.D., Abdellah, Z., Arrowsmith, C., Chillingworth, T., Churcher, C., et al. (2008). Insights from the complete genome sequence of *Mycobacterium marinum* on the evolution of *Mycobacterium tuberculosis*. *Genome Res.* 18, 729–741.
- Swaim, L.E., Connolly, L.E., Volkman, H.E., Humbert, O., Born, D.E., and Ramakrishnan, L. (2006). *Mycobacterium marinum* infection of adult zebrafish causes caseating granulomatous tuberculosis and is moderated by adaptive immunity. *Infect. Immun.* 74, 6108–6117.
- Szumowski, J.D., Adams, K.N., Edelstein, P.H., and Ramakrishnan, L. (2013). Antimicrobial efflux pumps and *Mycobacterium tuberculosis* drug tolerance: evolutionary considerations. *Curr. Top. Microbiol. Immunol.* 374, 81–108.
- Takaki, K., Davis, J.M., Winglee, K., and Ramakrishnan, L. (2013). Evaluation of the pathogenesis and treatment of *Mycobacterium marinum* infection in zebrafish. *Nat. Protoc.* 8, 1114–1124.
- Taylor, J.L., Hattle, J.M., Dreitz, S.A., Troudt, J.M., Izzo, L.S., Basaraba, R.J., Orme, I.M., Matrisian, L.M., and Izzo, A.A. (2006). Role for matrix metalloproteinase 9 in granuloma formation during pulmonary *Mycobacterium tuberculosis* infection. *Infect. Immun.* 74, 6135–6144.
- Tobin, D.M., and Ramakrishnan, L. (2008). Comparative pathogenesis of *Mycobacterium marinum* and *Mycobacterium tuberculosis*. *Cell. Microbiol.* 10, 1027–1039.
- Tobin, D.M., Vary, J.C., Jr., Ray, J.P., Walsh, G.S., Dunstan, S.J., Bang, N.D., Hagge, D.A., Khadge, S., King, M.C., Hawn, T.R., et al. (2010). The *Ita4h* locus modulates susceptibility to mycobacterial infection in zebrafish and humans. *Cell* 140, 717–730.
- Tobin, D.M., Roca, F.J., Oh, S.F., McFarland, R., Vickery, T.W., Ray, J.P., Ko, D.C., Zou, Y., Bang, N.D., Chau, T.T., et al. (2012). Host genotype-specific therapies can optimize the inflammatory response to mycobacterial infections. *Cell* 148, 434–446.
- Truman, R.W., Singh, P., Sharma, R., Busso, P., Rougemont, J., Paniz-Mondolfi, A., Kapopoulou, A., Brisse, S., Scollard, D.M., Gillis, T.P., and Cole, S.T. (2011). Probable zoonotic leprosy in the southern United States. *N. Engl. J. Med.* 364, 1626–1633.
- van der Wel, N., Hava, D., Houben, D., Fluitsma, D., van Zon, M., Pierson, J., Brenner, M., and Peters, P.J. (2007). *M. tuberculosis* and *M. leprae* translocate from the phagolysosome to the cytosol in myeloid cells. *Cell* 129, 1287–1298.
- Volkman, H.E., Clay, H., Beery, D., Chang, J.C.W., Sherman, D.R., and Ramakrishnan, L. (2004). Tuberculous granuloma formation is enhanced by a mycobacterium virulence determinant. *PLoS Biol.* 2, e367.
- Volkman, H.E., Pozos, T.C., Zheng, J., Davis, J.M., Rawls, J.F., and Ramakrishnan, L. (2010). Tuberculous granuloma induction via interaction of a bacterial secreted protein with host epithelium. *Science* 327, 466–469.
- Walburger, A., Koul, A., Ferrari, G., Nguyen, L., Prescianotto-Baschong, C., Huygen, K., Klebl, B., Thompson, C., Bacher, G., and Pieters, J. (2004). Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science* 304, 1800–1804.
- Wang, J., and Behr, M.A. (2014). Building a better bacillus: the emergence of *Mycobacterium tuberculosis*. *Front. Microbiol.* 5, 139.
- Watson, R.O., Manzanillo, P.S., and Cox, J.S. (2012). Extracellular *M. tuberculosis* DNA targets bacteria for autophagy by activating the host DNA-sensing pathway. *Cell* 150, 803–815.
- Wells, W.F., Ratcliffe, H.L., and Grumb, C. (1948). On the mechanics of droplet nuclei infection; quantitative experimental air-borne tuberculosis in rabbits. *Am. J. Hyg.* 47, 11–28.
- Yang, C.-T., Cambier, C.J., Davis, J.M., Hall, C.J., Crosier, P.S., and Ramakrishnan, L. (2012). Neutrophils exert protection in the early tuberculous granuloma by oxidative killing of mycobacteria phagocytosed from infected macrophages. *Cell Host Microbe* 12, 301–312.
- Zhang, Y.J., Reddy, M.C., Ioerger, T.R., Rothchild, A.C., Dartois, V., Schuster, B.M., Trauner, A., Wallis, D., Galaviz, S., Huttenhower, C., et al. (2013). Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. *Cell* 155, 1296–1308.
- Zumla, A., Raviglione, M., Hafner, R., and von Reyn, C.F. (2013). Tuberculosis. *N. Engl. J. Med.* 368, 745–755.

A Highly Conserved Program of Neuronal Microexons Is Misregulated in Autistic Brains

Manuel Irimia,^{1,2,*} Robert J. Weatheritt,^{1,3} Jonathan D. Ellis,¹ Neelroop N. Parikshak,⁴ Thomas Gonatopoulos-Pournatzis,¹ Mariana Babor,¹ Mathieu Quesnel-Vallières,¹ Javier Tapia,² Bushra Raj,¹ Dave O'Hanlon,¹ Miriam Barrios-Rodiles,⁶ Michael J.E. Sternberg,⁵ Sabine P. Cordes,^{6,7} Frederick P. Roth,^{1,6,7,8,9} Jeffrey L. Wrana,^{6,7} Daniel H. Geschwind,⁴ and Benjamin J. Blencowe^{1,7,*}

¹Donnelly Centre, University of Toronto, 160 College Street, Toronto, ON M5S 3E1, Canada

²EMBL/CRG Research Unit in Systems Biology, Centre for Genomic Regulation (CRG), 88 Dr. Aiguader, Barcelona 08003, Spain

³MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK

⁴Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, 695 Charles E. Young Drive South, Los Angeles, CA 90095, USA

⁵Centre for Integrative Systems Biology and Bioinformatics, Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

⁶Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, ON M5G 1X5, Canada

⁷Department of Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto, ON M5S 1A8, Canada

⁸Department of Computer Science, University of Toronto, 10 King's College Road, Toronto, ON M5S 3G4, Canada

⁹Canadian Institute For Advanced Research, 180 Dundas Street West, Toronto, ON M5G 1Z8, Canada

*Correspondence: mirimia@gmail.com (M.I.), b.blencowe@utoronto.ca (B.J.B.)

<http://dx.doi.org/10.1016/j.cell.2014.11.035>

SUMMARY

Alternative splicing (AS) generates vast transcriptomic and proteomic complexity. However, which of the myriad of detected AS events provide important biological functions is not well understood. Here, we define the largest program of functionally coordinated, neural-regulated AS described to date in mammals. Relative to all other types of AS within this program, 3–15 nucleotide “microexons” display the most striking evolutionary conservation and switch-like regulation. These microexons modulate the function of interaction domains of proteins involved in neurogenesis. Most neural microexons are regulated by the neuronal-specific splicing factor nSR100/SRRM4, through its binding to adjacent intronic enhancer motifs. Neural microexons are frequently misregulated in the brains of individuals with autism spectrum disorder, and this misregulation is associated with reduced levels of nSR100. The results thus reveal a highly conserved program of dynamic microexon regulation associated with the remodeling of protein-interaction networks during neurogenesis, the misregulation of which is linked to autism.

INTRODUCTION

Alternative splicing (AS)—the process by which different pairs of splice sites are selected in precursor mRNA to generate multiple mRNA and protein products—is responsible for greatly expanding the functional and regulatory capacity of metazoan genomes

(Braunschweig et al., 2013; Chen and Manley, 2009; Kalsotra and Cooper, 2011). For example, transcripts from over 95% of human multiexon genes undergo AS, and most of the resulting mRNA splice variants are variably expressed between different cell and tissue types (Pan et al., 2008; Wang et al., 2008). However, the function of the vast majority of AS events detected to date are not known, and new landscapes of AS regulation remain to be discovered and characterized (Braunschweig et al., 2014; Eom et al., 2013). Moreover, because the misregulation of AS frequently causes or contributes to human disease, there is a pressing need to systematically define the functions of splice variants in disease contexts.

AS generates transcriptomic complexity through differential selection of cassette alternative exons, alternative 5' and 3' splice sites, mutually exclusive exons, and alternative intron retention. These events are regulated by the interplay of *cis*-acting motifs and *trans*-acting factors that control the assembly of spliceosomes (Chen and Manley, 2009; Wahl et al., 2009). The assembly of spliceosomes at 5' and 3' splice sites is typically regulated by RNA-binding proteins (RBPs) that recognize proximal *cis*-elements, referred to as exonic/intronic splicing enhancers and silencers (Chen and Manley, 2009). An important advance that is facilitating a more general understanding of the role of individual AS events is the observation that many cell/tissue type- and developmentally-regulated AS events are coordinately controlled by individual RBPs, and that these events are significantly enriched in genes that operate in common biological processes and pathways (Calarco et al., 2011; Irimia and Blencowe, 2012; Licatalosi and Darnell, 2010).

AS can have dramatic consequences on protein function and/or affect the expression, localization, and stability of spliced mRNAs (Irimia and Blencowe, 2012). Whereas cell and tissue differentially regulated AS events are significantly underrepresented in functionally defined, folded domains in proteins, they

are enriched in regions of protein disorder that typically are surface accessible and embed short linear interaction motifs (Buljan et al., 2012; Ellis et al., 2012; Romero et al., 2006). AS events located in these regions are predicted to participate in interactions with proteins and other ligands (Buljan et al., 2012; Weatheritt et al., 2012). Indeed, among a set of analyzed neural-specific exons enriched in disordered regions, approximately one-third promoted or disrupted interactions with partner proteins (Ellis et al., 2012). These observations suggested that a widespread role for regulated exons is to specify cell and tissue type-specific protein-interaction networks.

Human disease mutations often disrupt *cis*-elements that control splicing and result in aberrant AS patterns (Cartegni et al., 2002). Other disease changes affect the activity or expression of RBPs, causing entire programs of AS to be misregulated. For example, amyotrophic lateral sclerosis-causing mutations in the RBPs TLS/FUS and TDP43 affect AS and other aspects of posttranscriptional regulation (Polymenidou et al., 2012), and changes in the expression of the RBP RBFOX1 have been linked to misregulation of AS in the brains of individuals with autism spectrum disorder (ASD) (Voineagu et al., 2011). It is also widely established that misregulation of AS plays important roles in altering the growth and invasiveness of various cancers (David and Manley, 2010). As is the case with assessing the normal functions of AS, it is generally not known which disease-misregulated AS events cause or contribute to disease phenotypes.

Central to addressing the above questions is the importance of comprehensively defining AS programs associated with normal and disease biology. Gene-prediction algorithms, high-throughput RNA sequencing (RNA-seq) analysis methods, and RNA-seq data sets generally lack the sensitivity and/or depth required to detect specific types of AS. In particular, microexons (Beachy et al., 1985; Coleman et al., 1987), defined here as 3–27 nucleotide (nt)-long exons, have been largely missed by genome annotations and transcriptome profiling studies (Volfovsky et al., 2003; Wu et al., 2013; Wu and Watanabe, 2005). This is especially true for microexons shorter than 15 nt. Furthermore, where alignment tools have been developed to capture microexons (Wu et al., 2013), they have not been applied to the analysis of different cell and tissue types or disease states.

In this study, we developed an RNA-seq pipeline for the systematic discovery and analysis of all classes of AS, including microexons. By applying this pipeline to deep RNA-seq data sets from more than 50 diverse cell and tissue types, as well as developmental stages, from human and mouse, we define a large program of neural-regulated AS. Strikingly, neural-included microexons represent the most highly conserved and dynamically regulated component of this program, and the corresponding genes are highly enriched in neuronal functions. These microexons are enriched on the surfaces of protein-interaction domains and are under strong selection pressure to preserve reading frame. We also observe that microexons are frequently misregulated in the brains of autistic individuals, and that this misregulation is linked to the reduced expression of the neural-specific Ser/Arg-related splicing factor of 100 kDa, nSR100/SRRM4. Collectively, our results reveal that alternative microexons represent the most highly conserved component of developmental AS regulation identified to date, and that they function in domain

surface “microsurgery” to control interaction networks associated with neurogenesis. Microexons thus represent a new landscape for investigating the molecular consequences of AS (mis) regulation in nervous system development and ASD.

RESULTS

Global Features of Neural-Regulated AS

An RNA-seq analysis pipeline was developed to detect and quantify all AS event classes involving all hypothetically possible splice junctions formed by the usage of annotated and unannotated splice sites, including those that demarcate microexons. By applying this pipeline to more than 50 diverse cell and tissue types, each from human and mouse (Table S1 available online), we identified ~2,500 neural-regulated AS events in each species (Figure 1A and Table S2; Extended Experimental Procedures).

Nearly half of the neural-regulated AS events, including alternative retained introns, are predicted to generate protein isoforms when the alternative sequence is both included and skipped. In contrast, only ~20% of AS events not subject to neural regulation (hereafter “non-neural” events) have the potential to generate alternative protein isoforms (Figure 1B; $p = 2.7 \times 10^{-248}$, proportion test). Gene Ontology (GO) analysis shows that genes with neural-regulated AS events predicted to generate alternative protein isoforms form highly interconnected networks based on functions associated with neuronal biology, signaling pathways, structural components of the cytoskeleton, and the plasma membrane (Figure 1C). Consistent with previous results (Fagnani et al., 2007; Pan et al., 2004), there is little overlap (8.5%) between genes with neural-regulated AS and mRNA expression, although these subsets of genes are highly enriched in overlapping GO terms (40% in common; Figure S1). These data reveal the largest program of neural-regulated AS events defined to date, and that this program is associated with a broader range of functional processes and pathways linked to nervous system biology than previously detected (Boutz et al., 2007; Fagnani et al., 2007; Ule et al., 2005).

Highly Conserved Microexons Are Frequently Neuron Specific

Further analysis of the neural-regulated AS program revealed a striking inverse relationship between the length of an alternative exon and its propensity to be specifically included in neural tissues. Increased neural-specific inclusion was detected for the majority of microexons (length ≤ 27 nt, Figure 2A); 60.7% of alternative microexons show increased neural “percent spliced in” (PSI) ($\Delta\text{PSI} > 15$) versus 9.5% of longer (average ~135 nt) alternative exons ($p = 1.9 \times 10^{-220}$, proportion test). This trend extends to microexons as short as 3 nt. RT-PCR validation experiments confirmed the RNA-seq-detected regulatory profiles and inclusion levels of all (10/10) microexons analyzed across ten diverse tissues ($R^2 = 0.92$, $n = 107$; Figure S2A). To further investigate the cell- and tissue-type specificity of microexon regulation, we used RNA-seq data (Sofueva et al., 2013; Zhang et al., 2013, 2014) to compare their inclusion levels in major glial cell types (astrocytes, microglia, and oligodendrocytes), in isolated neurons, and in muscle cells and tissues. Although up to ~20% of the detected neural-regulated microexons showed

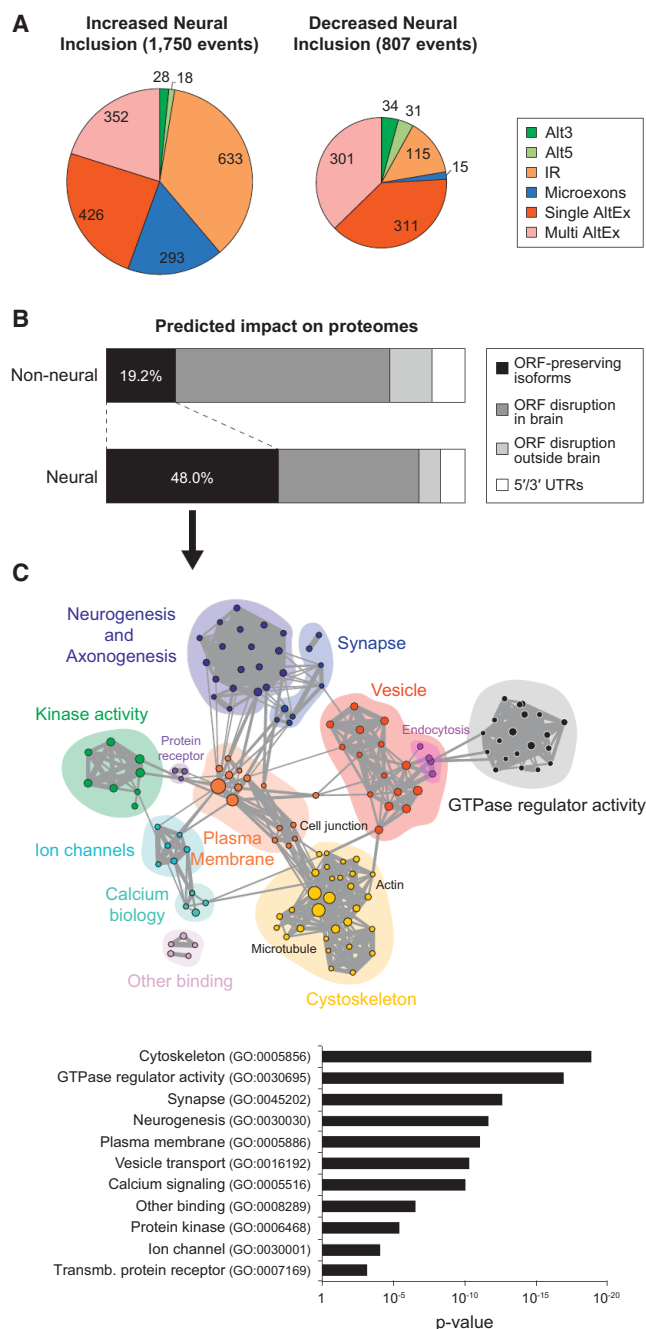


Figure 1. An Extensive Program of Neural-Regulated AS

(A) Distribution by type of human AS events with increased/decreased neural inclusion of the alternative sequence. Alt3/5, alternative splice-site acceptor/donor selection; IR, intron retention; Microexons, 3–27 nt exons; Single/Multi AltEx, single/multiple cassette exons.

(B) Predicted impact of non-neural and neural-regulated AS events on proteomes. Neural-regulated events are more often predicted to generate isoforms preserving open reading frame (ORF) when the alternative sequence is included and excluded (“ORF-preserving isoforms,” black), than to disrupt ORFs (i.e., the exon leads to a frameshift and/or introduces a premature termination codon) specifically in neural samples (“ORF disruption in brain,” dark gray) or in non-neural samples (“ORF preservation in brain,” light gray). See Extended Experimental Procedures for details.

increased PSIs in one or more glial cell types, and/or in muscle, compared to other non-neural tissues, the vast majority (>90%) of neural-regulated microexons displayed highest PSIs in neurons compared to all other cell and tissue types analyzed (Figures S2B–S2D and Extended Experimental Procedures). These results indicate that tissue-regulated microexons are predominantly neuronal specific.

Relative to longer alternative exons, microexons, in particular those that are 3–15 nt long and neural-specifically included, are strongly enriched in multiple features indicative of functionally important AS. They are highly enriched for lengths that are multiples of 3 nt (Figure 2B), and a significantly larger fraction are predicted to generate alternative protein isoforms upon inclusion and exclusion, compared with longer neural exons (Figure 2C; $p < 10^{-10}$, proportion test). They are also significantly more often conserved at the levels of genomic sequence, detection in alternatively spliced transcripts, and neural-differential regulation (Figures 2D and S2E, neural-regulated exons; $p < 0.001$ for all pairwise comparisons, proportion tests). Similar results were obtained when comparing neural-regulated microexons and longer exons that have matching distributions of neural versus non-neural Δ PSI values (data not shown). Of 308 neural-regulated microexons in human, 225 (73.5%) are neural-differentially spliced in mouse, compared to only 527 of 1,390 (37.9%) longer neural-regulated exons. Remarkably, although microexons represent only ~1% of all AS events, they comprise approximately one-third of all neural-regulated AS events conserved between human and mouse that are predicted to generate alternative protein isoforms (Figure S2F). Moreover, of ~150 analyzed mammalian, neural-regulated, 3–15 nt microexons, at least 55 are deeply conserved in vertebrate species spanning 400–450 million years of evolution, from zebrafish and/or shark to human (Table S3). This is in marked contrast to the generally low degree of evolutionary conservation of other types of AS across vertebrate species (Barbosa-Morais et al., 2012; Braunschweig et al., 2014; Merkin et al., 2012). Furthermore, comparable numbers of alternative microexons were detected in all analyzed vertebrate species, the majority of which are also strongly neural-specifically included (Figure 2E; Extended Experimental Procedures for details). Consistent with their striking regulatory conservation, sequences overlapping microexons, including both the upstream and downstream flanking intronic regions, are more highly conserved than sequences surrounding longer alternative exons (Figures 2F and S2G), including longer exons with a similar distribution of neural versus non-neural Δ PSI values (Figures S2H and S2I; data not shown).

Dynamic Regulation of Microexons during Neuronal Differentiation

To further investigate the functional significance of neural-regulated microexons, we used RNA-seq data to analyze their

(C) Enrichment map for GO and KEGG categories in genes with neural-regulated AS that are predicted to generate alternative protein isoforms (top) and representative GO terms and their associated enrichment p value for each subnetwork (bottom). The node size is proportional to the number of genes associated with the GO category and the width of the edges to the number of genes shared between GO categories.

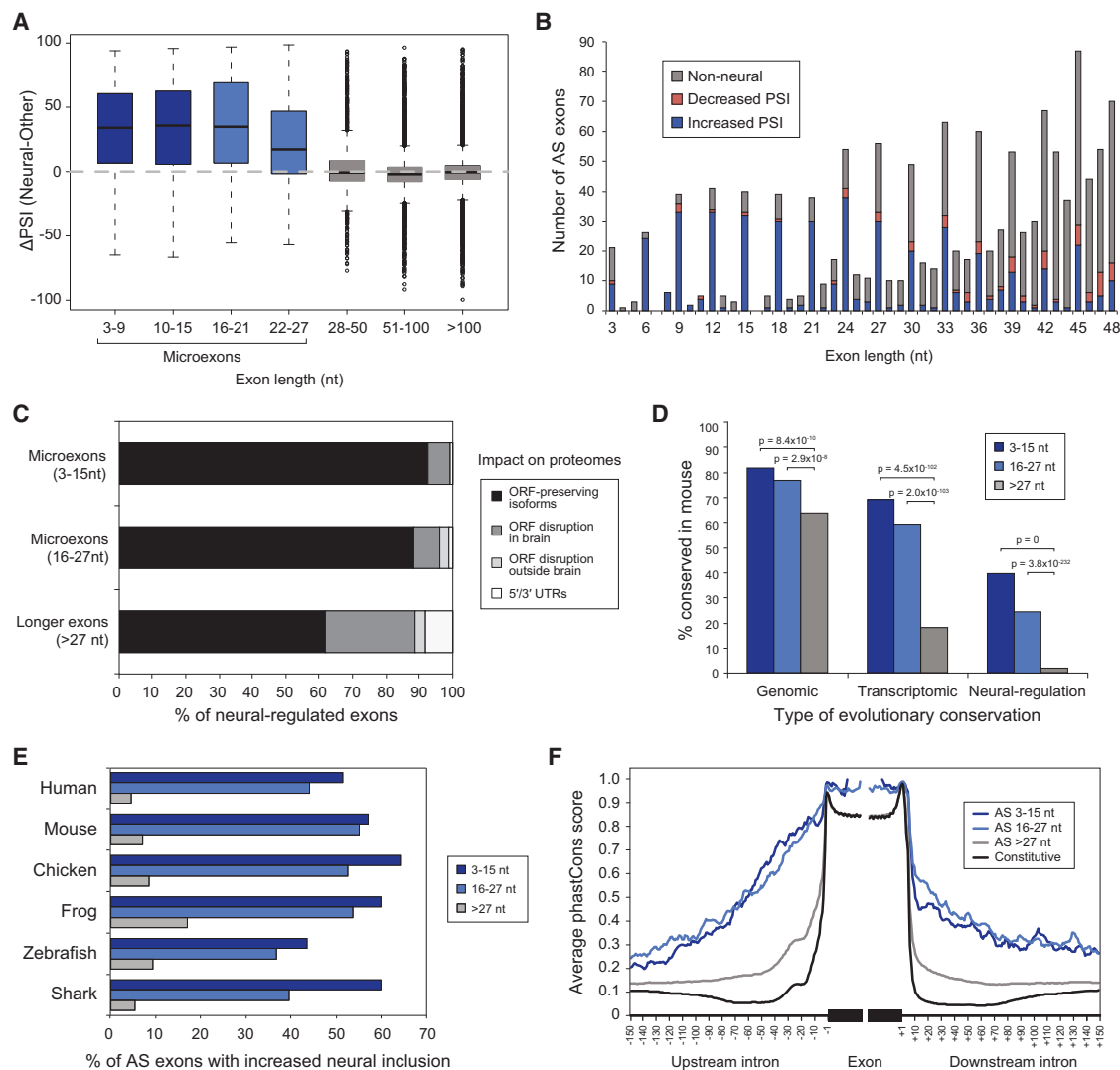


Figure 2. A Landscape of Highly Conserved Neural Microexons

(A) Difference in exon inclusion level (Δ PSI) between the average PSIs for neural samples and non-neural samples (y axis) for bins of increasing exon lengths (x axis). Microexons are defined as exons with lengths of 3–27 nt. Restricting the analysis to alternative exons with a PSI range across samples of >50 showed a similar pattern (data not shown).

(B) Number of exons by length whose inclusion levels are higher (blue), lower (red), or not different (gray) in neural compared to non-neural samples. Short exons tend to be multiple of 3 nt and have higher inclusion in neural samples.

(C) Percent of neural-regulated microexons (of lengths of 3–15 and 16–27 nt) and longer exons that are predicted to generate alternative ORF-preserving isoforms (black), disrupt the ORF in/outside neural tissues (dark/light gray), or overlap noncoding sequences (white).

(D) Higher evolutionary conservation of alternative microexons compared to longer alternative exons at the genomic, transcriptomic (i.e., whether the exon is alternatively spliced in both species), and neural-regulatory levels. y axis shows the percent of conservation at each specific level between human and mouse. p values correspond to two-sided proportion tests.

(E) Percent of alternative microexons and longer exons that are detected as neural-regulated (average absolute Δ PSI > 25) in each vertebrate species.

(F) Alternative 3–15 and 16–27 nt microexons show higher average phastCons scores at their intronic boundaries than longer alternative and constitutive exons. See also Figure S2.

regulation across six time points of differentiation of mouse embryonic stem cells (ESCs) into cortical glutamatergic neurons (Figure 3). Remarkably, of 219 neural-regulated microexons with sufficient read coverage across time points, 151 (69%) displayed a PSI switch ≥ 50 between ESCs and mature neurons, and 65 (30%) a switch of ≥ 90 (Figure 3). Unsupervised hierarchi-

cal clustering of PSI changes between consecutive time points (transitions T1 to T5) revealed several temporally distinct regulatory patterns (Figure 3A). Most microexons show sharp PSI switches at late (T3 to T5) transitions during differentiation. These stages correspond to maturing postmitotic neurons when pan-neuronal markers are already expressed and are subsequent

to the expression of most neurogenic transcription factors (Figure S3A). This pattern of late activation (Figure S3B) suggests enrichment for important functions for microexons in terminal neurogenesis (Figure 1C). Despite the small number of genes representing clusters of kinetically distinct sets of regulated microexons, each cluster revealed significant enrichment of specific GO terms including “regulation of GTPase activity” (Cluster I), “glutamate receptor binding,” and “actin cytoskeleton organization” (Cluster V) (Table S4). These observations indicate that the dynamic switch-like regulation of microexons is intimately associated with the maturation of neurons.

The Neural-Specific Splicing Factor nSR100/SRRM4 Regulates Most Neural Microexons

Among several analyzed splicing regulators (Extended Experimental Procedures), knockdown and overexpression of nSR100 had the strongest effect on microexon regulation, with more than half of the profiled microexons displaying a pronounced change in inclusion level compared to controls (Figures 4A and S4A–S4H). Moreover, an analysis of RNA-seq data from different neural cell types (Zhang et al., 2014) revealed that nSR100 has the strongest neuronal-specific expression relative to the other splicing regulators (Figure S4I and data not shown), which is also consistent with its immunohistochemical detection in neurons but not glia (Calarco et al., 2009). Recently, we have shown that nSR100 promotes the inclusion of a subset of (longer) neural exons via binding to intronic UGC motifs proximal to sub-optimal 3' splice sites (Raj et al., 2014). Consistent with these results, and supporting a direct role for nSR100 in microexon regulation, RNA sequence tags crosslinked to nSR100 *in vivo* are also highly enriched in intronic sequences containing UGC motifs, located adjacent to the 3' splice sites of nSR100-regulated microexons (Figures 4B and 4C; $p < 0.0001$ for all comparisons; Wilcoxon rank-sum test). We additionally observe that, relative to longer exons, neural-regulated microexons are associated with weak 3' splice sites and strong 5' splice sites (Figure S4J). nSR100 thus has a direct and extensive role in the regulation of the neural microexon program.

Distinct Protein-Regulatory Properties of Microexons

Neural-regulated microexons, in particular those that are 3–15 nt long, possess multiple properties that distinguish them from longer neural-regulated exons (Figures 5 and S5). A significantly smaller fraction overlap predicted disordered amino acid residues (Figures 5A and S5A–S5D; $p < 1.3 \times 10^{-4}$; three-way Fisher's exact tests), whereas a significantly higher fraction overlap modular protein domains (Figures 5B and S5E; ~2-fold increase, $p = 1.0 \times 10^{-54}$; proportion test). In contrast, microexon residues overlapping protein domains are significantly more often surface accessible and enriched in charged residues (Figures 5C, 5D, and S5F–S5I; $p < 10^{-7}$ for all comparisons; proportion test) than are residues overlapping longer neural or non-neural exons. Moreover, when not overlapping protein domains, microexons are significantly more often located immediately adjacent (i.e., within 5 amino acids) to folded protein domains (Figures 5E, S5J, and S5K). These results suggest that a common function of microexons may be to modulate the activity of overlapping or adjacent protein domains. Supporting this view,

among 49 available and modeled by homology tertiary protein structures containing microexons, the corresponding residues are largely surface accessible and unlikely to significantly affect the folding of the overlapping or adjacent protein domains (Figure S6A; Extended Experimental Procedures).

Microexons Modulate the Function of Interaction Domains

Neural-regulated microexons are significantly enriched in domains that function in peptide and lipid-binding interactions (Figures 5F and S5L; $p = 1.7 \times 10^{-6}$; proportion test). Overall, genes with microexons are highly enriched in modular domains involved in cellular signaling, such as SH3 and PH domains (Figure S5M). Conversely, unlike longer neural exons (Buljan et al., 2012; Ellis et al., 2012), they are depleted of linear binding motifs (Figures 5G and S5N; $p < 0.005$; proportion tests for all comparisons). Moreover, proteins containing microexons are significantly more often central in protein-protein interaction networks and detected in stable protein complexes compared to proteins with other types of alternative exons (Figures 5H, S5O, and S5P; $p \leq 0.004$ for all comparisons; Wilcoxon rank-sum test). Taken together with the data in Figure 1, these results suggest that microexons may often regulate interaction domains to facilitate the remodeling of protein-interaction networks associated with signaling and other aspects of neuronal maturation and function.

To test this hypothesis, we employed luminescence-based mammalian interactome mapping (LUMIER; Barrios-Rodiles et al., 2005; Ellis et al., 2012) and coimmunoprecipitation-western blot assays to investigate whether the insertion of a highly conserved, neural-regulated 6 nt microexon in the nuclear adaptor Apbb1 affects its known interactions with the histone acetyltransferase Kat5/Tip60 and amyloid precursor protein App (Figures 6A–6D). Previous genetic and functional studies have revealed multiple functions for the Apbb1-Kat5 complex (Cao and Sudhoff, 2001; Stante et al., 2009), and that the loss of Kat5 activity is associated with developmental defects that impact learning and memory (Pirooznia et al., 2012; Wang et al., 2004, 2009) (see Discussion). Apbb1 contains two phosphotyrosine-binding domains, PTB1 and PTB2, which bind Kat5 and App, respectively (Cao and Sudhoff, 2001). Exemplifying the distinct protein features of neural microexons described above (Figure 5), the Apbb1 microexon adds two charged residues (Arg and Glu) to the PTB1 domain near its predicted interaction surface (Figures 6A and 6B; Extended Experimental Procedures). LUMIER and coimmunoprecipitation-western analysis reveal that inclusion of the microexon significantly enhances the interaction with Kat5, whereas there is little to no effect on the interaction with App (Figures 6C, 6D, S6B, and S6C). Substitution of both microexon residues with alanine also enhanced the Kat5 interaction, although to a lesser extent than the presence of Arg and Glu (Figure 6C). This suggests that the primary function of this microexon is to extend the interface with which Apbb1 binds its partner proteins.

We also examined the function of a 9 nt microexon in the AP1S2 subunit of the adaptor-related protein complex 1 (AP1). The AP1 complex functions in the intracellular transport of cargo proteins between the *trans*-Golgi apparatus and endosomes by linking clathrin to the cargo proteins during vesicle membrane

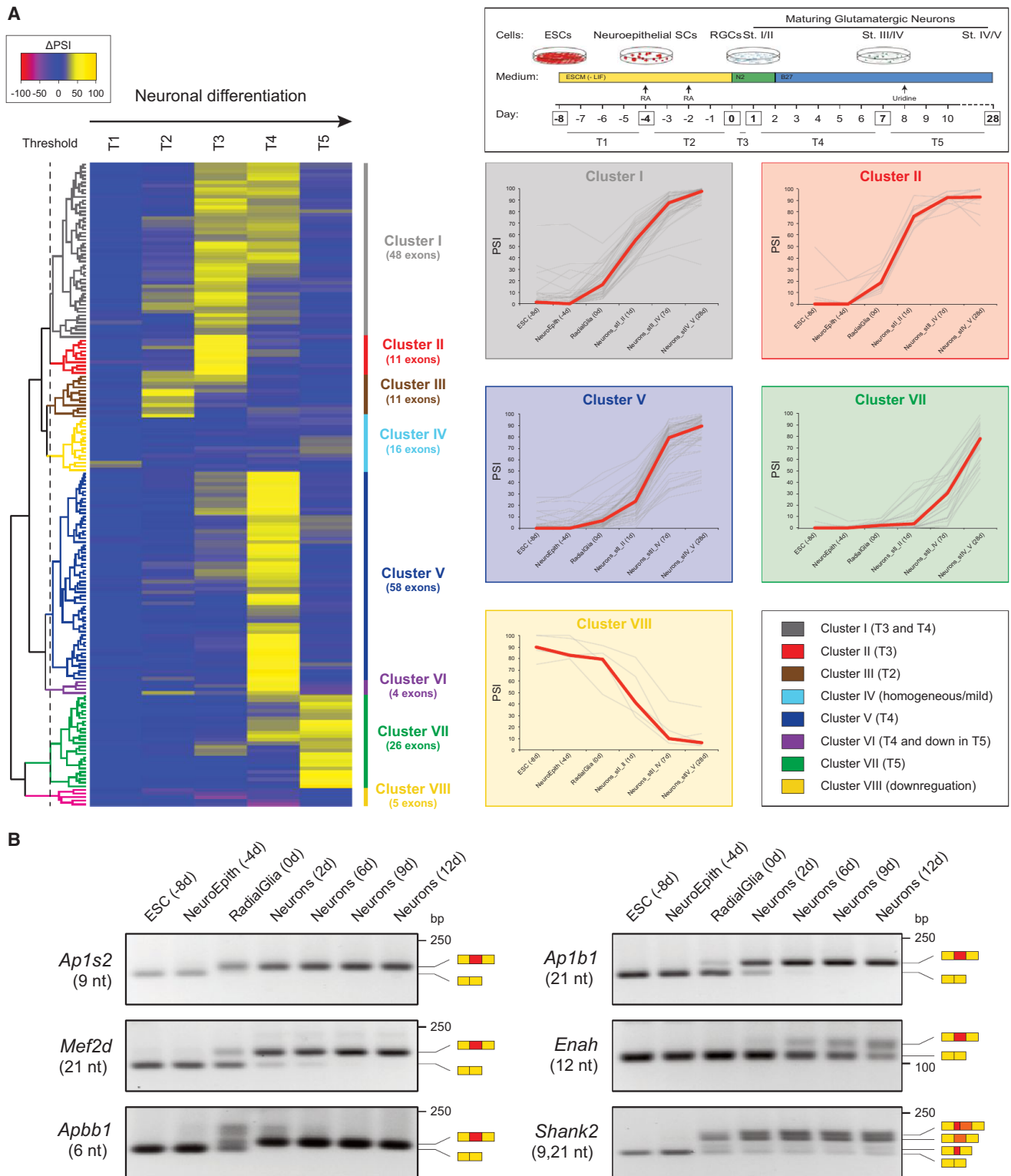


Figure 3. Switch-like Regulation of Microexons during Neuronal Differentiation

(A) Heatmap of PSI changes (Δ PSIs) between time points during differentiation of ESCs to glutamatergic neurons in vitro (Hubbard et al., 2013). Yellow/pink indicate increased/decreased PSI at a given transition (T1 to T5). Unsupervised clustering detects eight clusters of exons based on their dynamic PSI regulation

(legend continued on next page)

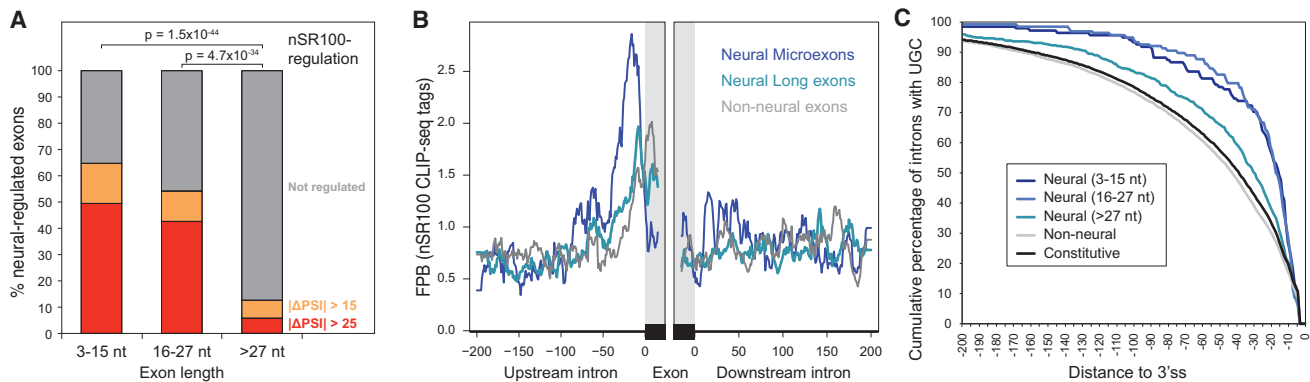


Figure 4. nSR100 Is a Positive, Direct Regulator of Most Microexons

(A) Percent of neural-regulated exons within each length class that are affected by nSR100 expression in human 293T kidney cells (absolute $\Delta\text{PSI} > 15$ [orange] or absolute $\Delta\text{PSI} > 25$ [red]). p values correspond to two-sided proportion tests of affected versus unaffected events.

(B) Average normalized density of nSR100-crosslinked sites in 200 nt windows encompassing neural-regulated exons of different length classes. FPB, fragments per billion.

(C) Cumulative distribution plots indicating the position of the first UGC motif within 200 nt upstream of neural-regulated microexons and longer exons, as well as non-neural and constitutive exons. $p < 0.0001$ for all comparisons against microexons, Wilcoxon rank-sum test.

See also Figure S4.

formation (Kirchhausen, 2000) and is important for the somato-dendritic transport of proteins required for neuronal polarity (Fariás et al., 2012). Interestingly, mutations in AP1S2 have been previously implicated in phenotypic features associated with ASD and X-linked mental retardation (Borck et al., 2008; Tarpey et al., 2006). Coimmunoprecipitation-western analyses reveal that the microexon in AP1S2 strongly promotes its interaction with another AP1 subunit, AP1B1 (Figures 6E and S6D). This observation thus provides additional evidence supporting an important role for microexons in the control of protein interactions that function in neurons.

Microexons Are Misregulated in Individuals with ASD

The properties of microexons described above suggest that their misregulation could be associated with neurological disorders. To investigate this possibility, we analyzed RNA-seq data from the superior temporal gyrus (Brodmann areas ba41/42/22) of postmortem samples from individuals with ASD and control subjects, matched for age, gender, and other variables (Experimental Procedures). These samples were stratified based on the strength of an ASD-associated gene-expression signature (Voineagu et al., 2011), and subsets of 12 ASD samples with the strongest ASD-associated differential gene-expression signatures and 12 controls were selected for further analysis. Remarkably, within these samples, 126 of 504 (30%) detected alternative microexons display a mean $\Delta\text{PSI} > 10$ between ASD and control subjects (Figure 7A); of these, 113 (90%) also display neural-differential regulation. By contrast, only 825 of 15,405 (5.4%) longer (i.e., >27 nt) exons show such misregulation

(Figure 7A); of these, 285 (35%) correspond to neural-regulated exons. Significant enrichment for misregulation among microexons compared to longer exons was also observed when restricting the analysis to neural-regulated exons, including subsets of neural-regulated microexons and longer exons with similar distributions of neural versus non-neural ΔPSI values (Figure S7A; $p < 2 \times 10^{-4}$; proportion test; data not shown). Similar results were observed when analyzing data from a different brain region (Brodmann area ba9) from the same individuals (data not shown). RT-PCR experiments on a representative subset of profiled tissues confirmed increased misregulation of microexons in autistic versus control brain samples (Figure S7B). Analysis of the proportions of microexons displaying coincident misregulation revealed that the vast majority (81.3%) have a $\Delta\text{PSI} > 10$ in at least half of the ASD-stratified brain samples (Figure S7C). However, only 26.9% (32/119) of the genes containing misregulated microexons overlapped with the 2,519 genes with significant ASD-associated misregulation at the level of gene expression. This reveals that largely distinct subsets of genes are misregulated at the levels of expression and microexon splicing in the analyzed ASD subjects. In contrast, a comparison of autistic subjects that possessed a weaker ASD-related differential gene-expression signature did not reveal significant misregulation of microexons or of longer exons (data not shown). These data reveal frequent misregulation of microexon splicing in the brain cortices of some individuals with ASD.

Consistent with a widespread and important role for nSR100 in the regulation of microexons (Figure 4), nSR100 mRNA

(clusters I–VIII, legend). Right, top: scheme of the neuronal differentiation assay, time points of sample collection and analyzed transitions. Right, bottom: PSIs for each microexons (gray lines) in five selected clusters; red lines show the median for the cluster at each time point.

(B) Representative RT-PCR assays monitoring AS patterns of microexons during neuronal differentiation in Ap1s2 (9 nt), Mef2d (21 nt), Apbb1 (6 nt), Ap1b1 (21 nt), Enah (12 nt), and Shank2 (9 and 21 nt).

See also Figure S3.

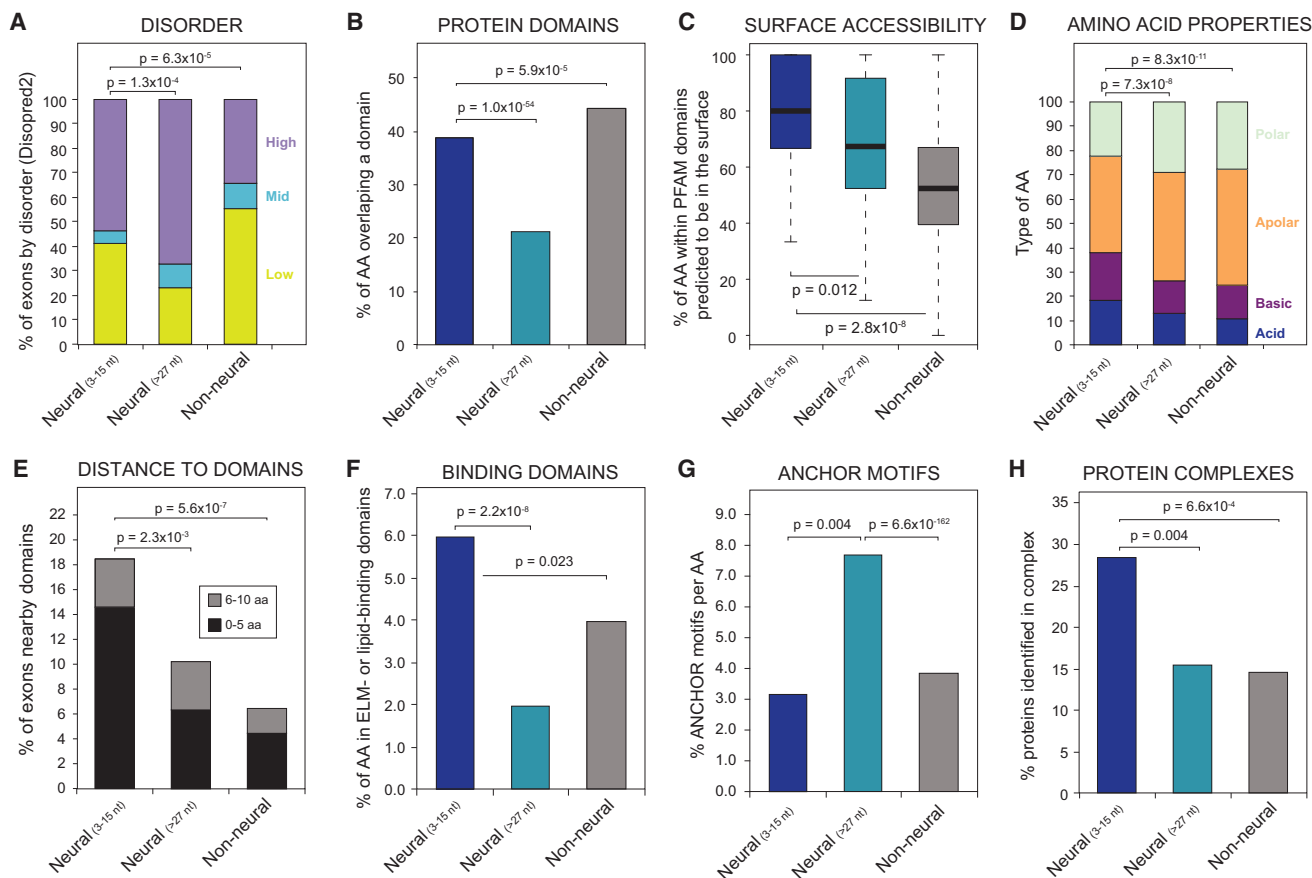


Figure 5. Microexons Possess Distinct Protein-Coding Features

For each analysis, values are shown for neural-regulated, 3–15 nt microexons and longer (>27 nt) exons, as well as non-neural AS exons (see Figure S5 for other types of exons).

(A) Percent of exons with a high average (>0.67), mid-range (0.33 to 0.67), and low disorder rate (<0.33).

(B) Fraction of amino acids (AA) that overlap a PFAM protein domain.

(C) Percent of AA within PFAM domains predicted to be on the protein surface.

(D) Percent of AA types based on their properties; p values correspond to the comparison of charged (acid and basic) versus uncharged (polar and apolar) AAs.

(E) Percent of exons that are adjacent to a domain (within 0–5 [black] or 6–10 AAs [gray]); p values correspond to the comparison of exons within 0–5 AAs.

(F) Percent of residues overlapping PFAM domains involved in linear motif or lipid binding.

(G) Percent of residues overlapping binding motifs predicted by ANCHOR.

(H) Percent of residues with proteins identified as belonging to one or more protein complexes (data from Havugimana et al., 2012).

All p values correspond to proportion tests except for (A) (three-way Fisher's test) and (C) (Wilcoxon rank-sum test). See also Figure S5.

expression is, on average, significantly downregulated in the brains of the analyzed ASD versus control subjects and to an even greater extent in brain samples with the strongest ASD-associated signature compared to the controls ($\sim 10\%$, $p = 0.014$, FDR < 0.1, Figure 7B and data not shown). These differences were confirmed by qRT-PCR assays for a representative subset of individuals ($p < 2.8 \times 10^{-4}$ for all normalizations; two-sided t test; Figure S7D). Moreover, relative to other exons, nSR100-dependent microexons are significantly more often misregulated in brain tissues from ASD compared to control subjects (Figure 7C; $p < 0.01$ for all comparisons; proportion test). Notably, we also observe significantly higher correlations between microexon inclusion and nSR100 mRNA expression levels across the stratified ASD samples and controls for those microexons regulated by nSR100 relative to those microexons that are

not regulated by this factor (Figure 7D; $p = 1.4 \times 10^{-7}$; Wilcoxon rank-sum test).

A GO analysis of genes with ASD-associated misregulation of microexons reveals significant enrichment of terms related to axonogenesis and synapse biology (Figure 7E), processes that have been previously implicated in autism (Gilman et al., 2011; Parikshak et al., 2013; Voineagu et al., 2011). Many of the corresponding genes act in common pathways and/or physically interact through protein-protein interactions (Figure 7F). Moreover, misregulated microexons are also significantly enriched in genes that have been genetically linked to ASD ($p < 0.0005$; Fisher's exact test), including many relatively well-established examples such as *DNTA*, *ANK2*, *ROBO1*, *SHANK2*, and *AP1S2*. Other genes with misregulated microexons have been linked to learning or intellectual disability (e.g., *APBB1*,

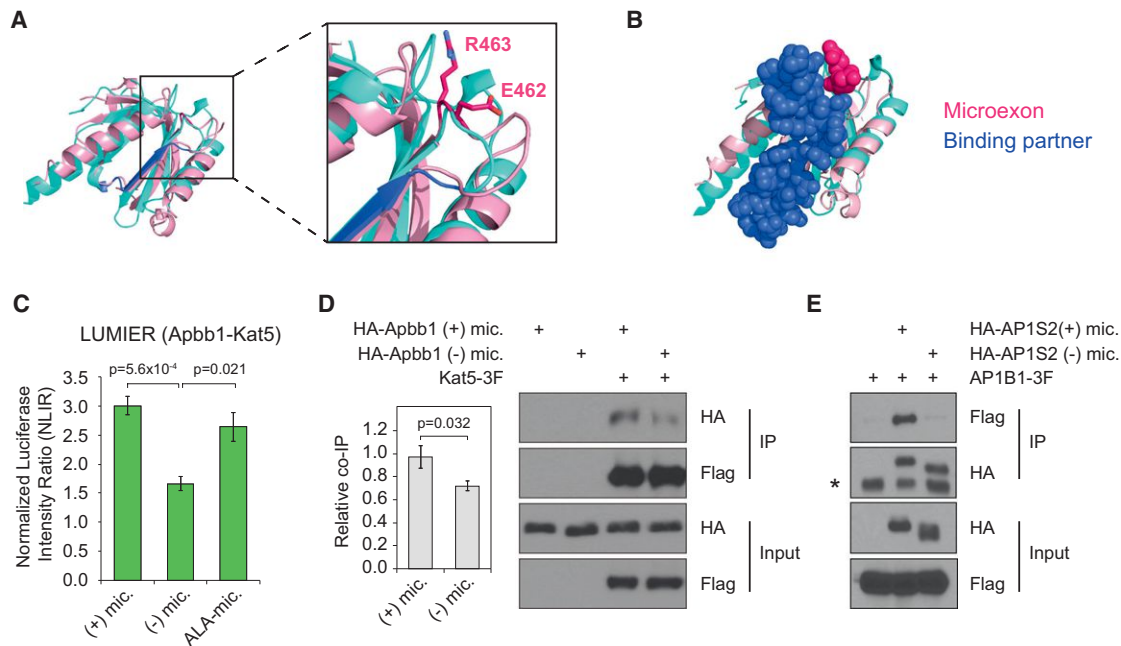


Figure 6. Microexons Regulate Protein-Protein Interactions

(A) Structural alignment of APBB1-PTB1 (pink) and APBB1-PTB2 (cyan) domains. Residues located at the protein-binding interface of APBB1-PTB2 are shown in blue. Inset shows the microexon residues in APBB1-PTB1 (E462-R463).

(B) Upon superimposition of APBB1-PTB1 (pink) and APBB1-PTB2 (cyan) domains, the microexon (magenta) is located close to the APBB1-PTB2-binding partner (APP protein fragment, blue), suggesting that the microexon in PTB1 may affect protein binding.

(C) Quantification of LUMIER-normalized luciferase intensity ratio (NLIR) values for RL-tagged Apbb1, with or without the microexon, or with a mutated version consisting of two alanine substitutions (ALA-mic.), coimmunoprecipitated with 3Flag-tagged Kat5.

(D and E) 293T cells were transfected HA-tagged Apbb1 (D) or AP1S2 (E) constructs, with or without the respective microexon, together with 3Flag-tagged Kat5 (D) or AP1B1 (E), as indicated. Immunoprecipitation was performed with anti-Flag (D) or anti-HA (E) antibody, and the immunoprecipitates were blotted with anti-HA or anti-Flag antibody, as indicated. Results shown in (E) were confirmed in a biological replicate experiment (Figure S6D).

p values in (C) and (D) correspond to t tests for four and three replicates, respectively; error bars indicate SEM. Asterisk in (E) indicates a band corresponding to the light chain of the HA antibody.

TRAPPC9, and *RAB3GAP1*). In this regard, it is interesting to note that the microexons we have analyzed in APBB1 and AP1S2 are significantly misregulated in the brain samples from ASD subjects ($p < 0.05$; Wilcoxon rank-sum test; Figure S7E). Taken together with data in Figures 5 and 6, the results suggest that the misregulation of microexons, as well as of longer alternative exons (Corominas et al., 2014; Voineagu et al., 2011), may impact protein-interaction networks that are required for normal neuronal development and synaptic function. Disruption of microexon-regulated protein-interaction networks is therefore a potentially important mechanism underlying ASD and likely other neurodevelopmental disorders.

DISCUSSION

In this study, we show that alternative microexons display the highest degrees of genomic sequence conservation, tissue-specific regulatory conservation, and frame-preservation potential, relative to all other classes of AS detected to date in vertebrate species. Unlike longer neural-regulated exons, neural microexons are significantly enriched in surface-accessible, charged amino acids that overlap or lie in close proximity to protein domains, including those that bind linear motifs. Together with their

remarkably dynamic regulation, these observations suggest that microexons contribute important and complementary roles to longer neural exons in the remodeling of protein-interaction networks that operate during neuronal maturation.

Most microexons display high inclusion at late stages of neuronal differentiation in genes (e.g., *Src* [Black, 1991], *Bin1*, *Agrn*, *Dock9*, *Shank2*, and *Robo1*) associated with axonogenesis and the formation and function of synapses. Supporting such functions, an alternative microexon overlapping the SH3A domain of Intersectin 1 (*Itsn1*) has been reported to promote an interaction with Dynamin 1 and was proposed to modulate roles of *Itsn1* in endocytosis, cell signaling, and/or actin-cytoskeleton dynamics (Dergai et al., 2010). A neural-specific microexon in Protrudin/Zfyve27 was recently shown to increase its interaction with the vesicle-associated membrane protein-associated protein (VAP) and to promote neurite outgrowth (Ohnishi et al., 2014). Similarly, in the present study, we show that a 6 nt neural microexon in Apbb1/Fe65 promotes an interaction with Kat5/Tip60. Apbb1 is an adaptor protein that functions in neurite outgrowth (Cheung et al., 2014; Ikin et al., 2007) and synaptic plasticity (Sabo et al., 2003), processes that have been linked to neurological disorders including ASD (Hussman et al., 2011). Consistent with these findings, we have previously shown

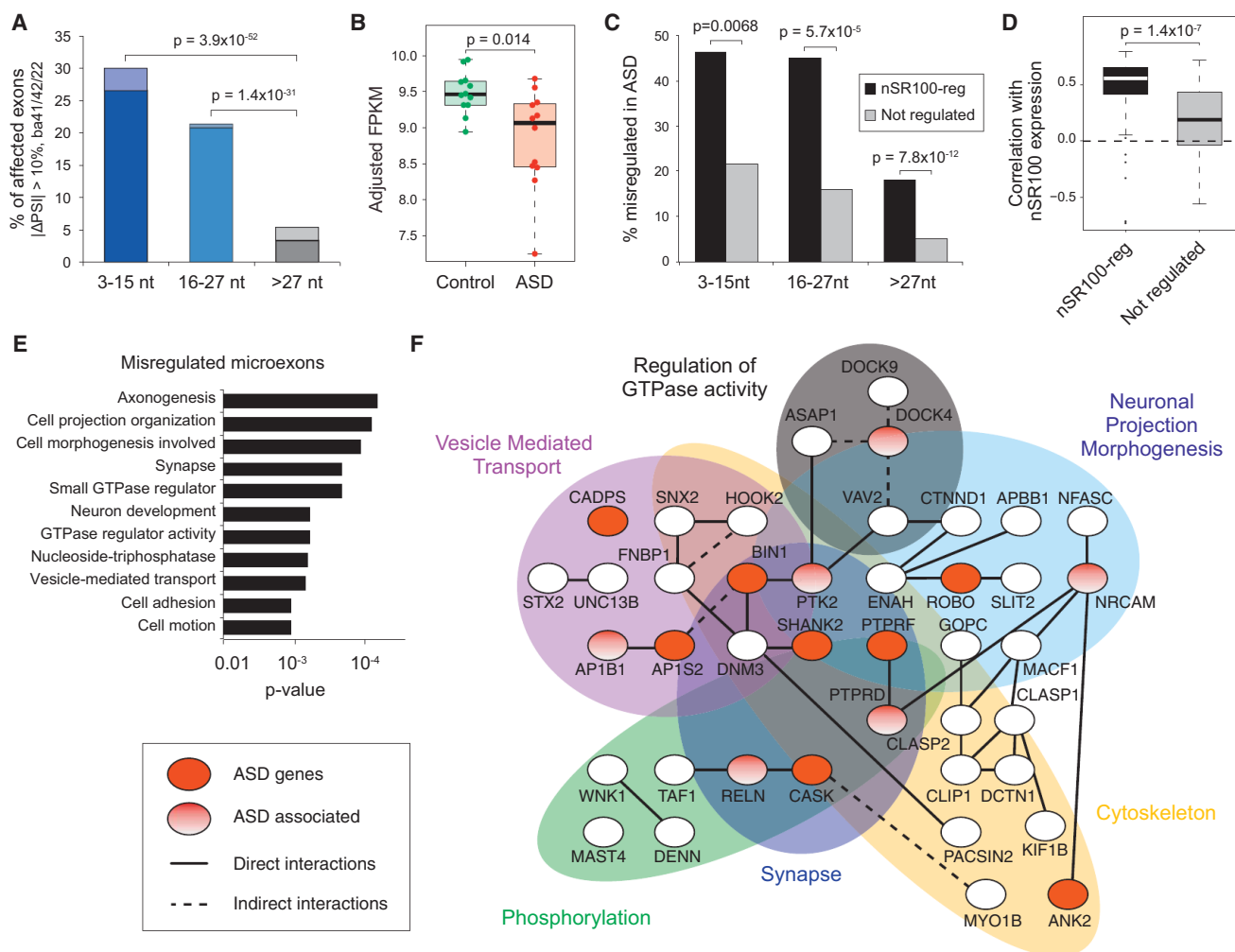


Figure 7. Microexons Are Often Misregulated in ASD

(A) Percent of alternative exons of each length class that are misregulated in ASD (absolute $\Delta\text{PSI} > 10$ between PSI-averaged ASD and control groups in ba41/42/22 brain regions). Dark shading, lower inclusion in ASD; light shading, higher inclusion in ASD; p values correspond to proportion tests.

(B) Expression of nSR100 across the 12 control and 12 ASD individuals. Adjusted FPKMs were calculated using a regression analysis that accounts for variation derived from differences in RNA integrity, brain sample batch, sequencing depth, and 5'-3' bias in measurements of gene-level FPKM values.

(C) Percent of exons within each length class misregulated in autistic compared to control brains (average absolute $\Delta\text{PSI} > 10$) for nSR100-regulated ($\Delta\text{PSI} > 25$ in the nSR100-overexpressing compared to control 293T cells) and non-nSR100-regulated (absolute $\Delta\text{PSI} < 5$) exons.

(D) Distribution of correlation coefficients between PSIs and nSR100 expression values across stratified ASD and control samples for microexons that are (n = 59) or are not (n = 69) regulated by nSR100. Only microexons with sufficient read coverage to derive accurate PSI quantifications in at least 9 ASD and 9 control ba41/42/22 samples were included. p value correspond to Wilcoxon rank-sum test.

(E) GO categories significantly enriched in genes with microexons that are misregulated in ASD.

(F) A protein-protein interaction network involving genes with ASD misregulated microexons ($\Delta\text{PSI} > 10$) in ba41/42/22 brain regions. Genes with major effect mutations, and smaller effect risk genes, are indicated in red and shaded ovals, respectively. Genes grouped by functional category are indicated.

See also Figure S7.

that nSR100 promotes neurite outgrowth (Calarco et al., 2009). In the present study, we further demonstrate that it controls the switch-like regulation of most neural microexons, and that its reduced expression is linked to the altered splicing of microexons in the brains of subjects with ASD.

Many of the conserved, neural-regulated microexons identified in this study are misregulated in ASD individuals, including the microexon in AP1S2 that strongly promotes an interaction with the AP1B1 subunit of the AP1 intracellular transport com-

plex. Intriguingly, several other genes containing microexons are genetically linked to ASD, intellectual disability, and/or functions in memory and learning (see Results). Another link to ASD is the observation that nSR100 is strongly coexpressed in the developing human brain in a gene network module, M2, which is enriched for rare de novo ASD-associated mutations (Parikshak et al., 2013). Furthermore, additional genes containing microexons may have as yet undiscovered roles in ASD and or other neuropsychiatric disorders. For example, the microexon

in APBB1 is also significantly misregulated in brain tissues from ASD subjects (Figures S7B and S7E). It is possible that the misregulation of microexons, at least in part through altered expression of nSR100, perturbs protein-interaction networks required for proper neuronal maturation and function, thus contributing to ASD as well as other neurodevelopmental disorders. Consistent with this view, recent reports have begun to link individual microexons with neurodevelopmental disorders, including ASD (Zhu et al., 2014), schizophrenia (Ovadia and Shifman, 2011), and epilepsy (Rusconi et al., 2014). The discovery and characterization of widespread, neural-regulated microexons in the present study thus enable a systematic investigation of new and highly conserved mechanisms controlling protein-interaction networks associated with vertebrate nervous system development and neurological disorders.

EXPERIMENTAL PROCEDURES

RNA-Seq Data and Genomes

Unless stated otherwise, RNA-seq data were generated from poly(A)⁺ RNA (Table S1). Analyses used the following genome releases: *Homo sapiens*, hg19; *Mus musculus*, mm9; *Gallus gallus*, galGal3; *Xenopus tropicalis*, xenTro3; *Danio rerio*, danRer7; *Callorhinchus milii*, v1.0.

AS Analysis Pipeline

A multimodule analysis pipeline was developed that uses RNA-seq, expressed sequence tag (EST), and cDNA data, as well as gene annotations and evolutionary conservation, to assemble libraries of exon-exon junctions (EEJs) for subsequent read alignment to detect and quantify AS events in RNA-seq data. For cassette exons, three complementary modules were developed for assembling EEJs: (1) a “transcript-based module,” employing cufflinks (Trapnell et al., 2010) and alignments of ESTs and cDNAs with genomic sequence (Khare et al., 2012); (2) a “splice site-based module,” utilizing joining of all hypothetically possible EEJ combinations from annotated and de novo splice sites (Han et al., 2013); and (3) a “microexon module,” including de novo searching of pairs of donor and acceptor splice sites in intronic sequence. Alt3 or Alt5 events were quantified based on the fraction of reads supporting the usage of each alternative splice site. Intron retention was analyzed as recently described (Braunschweig et al., 2014). See Extended Experimental Procedures for additional details. All described human microexons and associated features are provided in Tables S5 and S6.

LUMIER Assay

HEK293T cells were transiently transfected using Polyfect (QIAGEN) with *Renilla* luciferase (RL)-tagged Apbb1, with or without inclusion of the microexon, or with a version consisting of two alanine substitutions, together with 3Flag-tagged Kat5. Subsequent steps were performed essentially as described previously (Ellis et al., 2012).

Immunoprecipitation and Immunoblotting

HEK293T cells were transiently transfected using Lipofectamine 2000 (Life Technologies). Cells were lysed in 0.5% TNTE. After preclearing with protein G-Sepharose, lysates were incubated with anti-Flag M2 antibody (Sigma) or anti-Hemagglutinin (HA) antibody (Roche) bound to Protein-G Dynabeads (Life Technologies) for 2 hr at 4°C. Immunoprecipitates were washed five times with 0.1% TNTE, subjected to SDS-PAGE, transferred onto nitrocellulose, and immunoblotted with the anti-HA antibody (Roche) or anti-Flag M2 antibody (Sigma). Detection was achieved using horseradish peroxidase-conjugated rabbit anti-rat (Sigma) or sheep anti-mouse secondary antibodies (GE Healthcare) and chemiluminescence. ImageJ was used for quantification of band intensities.

Analysis of Microexon Regulation

Available RNA-seq data from splicing factor-deficient or -overexpressing systems were used to identify misregulated exons and microexons (see Extended

Experimental Procedures). To investigate regulation by nSR100, we used PAR-iCLIP data and motif enrichments analyses, as recently described (Raj et al., 2014).

Comparison of ASD and Control Brain Samples

We analyzed 22 autistic individuals and 20 controls matched by age and gender. Samples from superior temporal gyrus (Brodmann areas ba41/42/22) were dissected, retaining gray matter from all cortical layers, and RNA was isolated using the miRNeasy kit (QIAGEN). Ribosomal RNA was depleted from 2 µg total RNA with the Ribo-Zero Gold kit (Epicenter) and then size-selected with AMPure XP beads (Beckman Coulter). An average of 64 million, 50 bp paired-end reads were generated for each sample (Table S1). The 12 case and 12 control samples with the strongest ASD-associated differential gene-expression signature were selected for downstream analyses (Extended Experimental Procedures for details). Sample selection was independent of any information on splicing changes.

ACCESSION NUMBERS

The BioProject ID for the RNA-seq data reported in this paper is PRJNA268211. The Gene Expression Omnibus (GEO) accession number for the RNA-seq data is GSE64018.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.035>.

AUTHOR CONTRIBUTIONS

M.I. developed the RNA-seq analysis pipeline and performed analyses in Figures 1, 2, 3, 4, 5, and 7. R.J.W., J.E., and N.N.P. contributed equally to this study, performing analyses of microexon protein sequence features (Figure 5), protein-interaction experiments (Figure 6), and analyses of autism patient RNA-seq data (Figure 7), respectively. T.G.-P. performed neuronal differentiation of ESCs and RT-PCR assays. M.Q.-V. performed RT-PCR assays. M.B. and J.T. analyzed and modeled protein structural data. B.R. generated RNA-seq datasets. D.O'H. assisted with cloning and protein-interaction assays. M.B.-R. optimized LUMIER assays. B.J.B., M.I., M.J.E.S., S.P.C., F.P.R., J.L.W., and D.H.G. supervised experiments and analyses. M.I. and B.J.B. designed the study and wrote the paper, with input from the other authors.

ACKNOWLEDGMENTS

The authors thank the Eunice Kennedy Shriver NICHD Brain and Tissue Bank for Developmental Disorders, the Autism Tissue Program, and the Harvard Brain Tissue Resource Center for providing brain samples. Dax Torti and Danica Leung of the Donnelly Sequencing Centre are gratefully acknowledged for sequencing samples. The authors also thank Xinchun Wang for initial contributions to the RNA-seq analysis pipeline, Ulrich Braunschweig for assistance with CLIP-seq analyses, Benjamin Lang for advice on surface accessibility measurements, Nuno Barbosa-Morais for guidance on statistical testing, and Serge Gueroussov and Jonathan Roth for helpful discussions and comments on the manuscript. M.I. holds an LTF from the Human Frontiers Science Program Organization. R.J.W. holds a Canadian Institute of Health Research (CIHR) Postdoctoral Fellowship. N.N.P. holds an NIMH NRSA fellowship. M.B. is supported by a fellowship from the Department of Cell and Systems Biology, University of Toronto. M.Q.-V. holds a Banting and Best CIHR Scholarship. T.G.-P. is supported by fellowships from EMBO and OSCI. This research was supported by grants from the CIHR (B.J.B., J.L.W., S.P.C.), Ontario Research Fund (J.L.W., B.J.B., and others), Alzheimer's Society, Canada (B.J.B.), University of Toronto McLaughlin Centre (B.J.B.), NIH/NHGRI (P50 HG004233 and U01HG001715) (F.P.R.), the Krembil Foundation (F.P.R.), the Avon Foundation (F.P.R.), NIMH (5R37MH060233 and 5R01MH094714) (D.H.G.), and the Simons Foundation (SFARI 206744) (D.H.G.). F.P.R. was also supported by the Canada Excellence Research

Chairs Program. B.J.B. holds the Banbury Chair of Medical Research at the University of Toronto.

Received: August 7, 2014

Revised: October 20, 2014

Accepted: November 18, 2014

Published: December 18, 2014

REFERENCES

- Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodenic, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587–1593.
- Barrios-Rodiles, M., Brown, K.R., Ozdamar, B., Bose, R., Liu, Z., Donovan, R.S., Shinjo, F., Liu, Y., Dembowy, J., Taylor, I.W., et al. (2005). High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* 307, 1621–1625.
- Beachy, P.A., Helfand, S.L., and Hogness, D.S. (1985). Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature* 313, 545–551.
- Black, D.L. (1991). Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? *Genes Dev.* 5, 389–402.
- Borck, G., Mollà-Herman, A., Boddaert, N., Encha-Razavi, F., Philippe, A., Robel, L., Desguerre, I., Brunelle, F., Benmerah, A., Munnich, A., and Colleaux, L. (2008). Clinical, cellular, and neuropathological consequences of AP1S2 mutations: further delineation of a recognizable X-linked mental retardation syndrome. *Hum. Mutat.* 29, 966–974.
- Boutz, P.L., Stoilov, P., Li, Q., Lin, C.H., Chawla, G., Ostrow, K., Shiue, L., Ares, M.J., Jr., and Black, D.L. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* 21, 1636–1652.
- Braunschweig, U., Gueroussov, S., Plocik, A.M., Graveley, B.R., and Blencowe, B.J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell* 152, 1252–1269.
- Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., and Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcripts. *Genome Res.* 24, 1774–1786.
- Buljan, M., Chalancon, G., Eustermann, S., Wagner, G.P., Fuxreiter, M., Bateman, A., and Babu, M.M. (2012). Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell* 46, 871–883.
- Calarco, J.A., Superina, S., O'Hanlon, D., Gabut, M., Raj, B., Pan, Q., Skalska, U., Clarke, L., Gelinis, D., van der Kooy, D., et al. (2009). Regulation of vertebrate nervous system alternative splicing and development by an SR-related protein. *Cell* 138, 898–910.
- Calarco, J.A., Zhen, M., and Blencowe, B.J. (2011). Networking in a global world: establishing functional connections between neural splicing regulators and their target transcripts. *RNA* 17, 775–791.
- Cao, X., and Sudhoff, T.C. (2001). A transcriptionally active complex of APP with Fe65 and histone acetyltransferase Tip60. *Science* 293, 115–120.
- Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 3, 285–298.
- Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* 10, 741–754.
- Cheung, H.N., Dunbar, C., Mórotz, G.M., Cheng, W.H., Chan, H.Y., Miller, C.C., and Lau, K.F. (2014). FE65 interacts with ADP-ribosylation factor 6 to promote neurite outgrowth. *FASEB J.* 28, 337–349.
- Coleman, K.G., Poole, S.J., Weir, M.P., Soeller, W.C., and Kornberg, T. (1987). The inverted gene of *Drosophila*: sequence analysis and expression studies reveal a close kinship to the engrailed gene. *Genes Dev.* 1, 19–28.
- Corominas, R., Yang, X., Lin, G.N., Kang, S., Shen, Y., Ghamsari, L., Broly, M., Rodriguez, M., Tam, S., Trigg, S.A., et al. (2014). Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat. Commun.* 5, 3650.
- David, C.J., and Manley, J.L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes Dev.* 24, 2343–2364.
- Dergai, M., Tsyba, L., Dergai, O., Zlatskii, I., Skrypkina, I., Kovalenko, V., and Rynditch, A. (2010). Microexon-based regulation of ITSN1 and Src SH3 domains specificity relies on introduction of charged amino acids into the interaction interface. *Biochem. Biophys. Res. Commun.* 399, 307–312.
- Ellis, J.D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J.A., Wang, X., Pan, Q., O'Hanlon, D., Kim, P.M., et al. (2012). Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* 46, 884–892.
- Eom, T., Zhang, C., Wang, H., Lay, K., Fak, J., Noebels, J.L., and Darnell, R.B. (2013). NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. *Elife* 2, e00178.
- Fagnani, M., Barash, Y., Ip, J.Y., Misquitta, C., Pan, Q., Saltzman, A.L., Shai, O., Lee, L., Rozenhek, A., Mohammad, N., et al. (2007). Functional coordination of alternative splicing in the mammalian central nervous system. *Genome Biol.* 8, R108.
- Fariás, G.G., Cuitino, L., Guo, X., Ren, X., Jarnik, M., Mattera, R., and Bonifacino, J.S. (2012). Signal-mediated, AP-1/clathrin-dependent sorting of transmembrane receptors to the somatodendritic domain of hippocampal neurons. *Neuron* 75, 810–823.
- Gilman, S.R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., and Vitkup, D. (2011). Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* 70, 898–907.
- Han, H., Irimia, M., Ross, P.J., Sung, H.K., Alipanahi, B., David, L., Golipour, A., Gabut, M., Michael, I.P., Nachman, E.N., et al. (2013). MBNL proteins repress ES-cell-specific alternative splicing and reprogramming. *Nature* 498, 241–245.
- Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phansie, S., et al. (2012). A census of human soluble protein complexes. *Cell* 150, 1068–1081.
- Hubbard, K.S., Gut, I.M., Lyman, M.E., and McNutt, P.M. (2013). Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs. *F1000Res* 2, 35. <http://dx.doi.org/10.12688/f1000research.2-35.v1>.
- Hussman, J.P., Chung, R.H., Griswold, A.J., Jaworski, J.M., Salyakina, D., Ma, D., Konidari, I., Whitehead, P.L., Vance, J.M., Martin, E.R., et al. (2011). A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. *Mol. Autism* 2, 1.
- Ikin, A.F., Sabo, S.L., Lanier, L.M., and Buxbaum, J.D. (2007). A macromolecular complex involving the amyloid precursor protein (APP) and the cytosolic adapter FE65 is a negative regulator of axon branching. *Mol. Cell. Neurosci.* 35, 57–63.
- Irimia, M., and Blencowe, B.J. (2012). Alternative splicing: decoding an expansive regulatory layer. *Curr. Opin. Cell Biol.* 24, 323–332.
- Kalsotra, A., and Cooper, T.A. (2011). Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* 12, 715–729.
- Khare, T., Pai, S., Koncencius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., Irimia, M., Jia, P., Ptak, C., Xia, M., et al. (2012). 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat. Struct. Mol. Biol.* 19, 1037–1043.
- Kirchhausen, T. (2000). Clathrin. *Annu. Rev. Biochem.* 69, 699–727.
- Licatalosi, D.D., and Darnell, R.B. (2010). RNA processing and its regulation: global insights into biological networks. *Nat. Rev. Genet.* 11, 75–87.

- Merkin, J., Russell, C.B., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* 338, 1593–1599.
- Ohnishi, T., Shirane, M., Hashimoto, Y., Saita, S., and Nakayama, K.I. (2014). Identification and characterization of a neuron-specific isoform of protrudin. *Genes Cells* 19, 97–111.
- Ovadia, G., and Shifman, S. (2011). The genetic variation of RELN expression in schizophrenia and bipolar disorder. *PLoS ONE* 6, e19955.
- Pan, Q., Shai, O., Misquitta, C., Zhang, W., Saltzman, A.L., Mohammad, N., Babak, T., Siu, H., Hughes, T.R., Morris, Q.D., et al. (2004). Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* 16, 929–941.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415.
- Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 155, 1008–1021.
- Pirooznia, S.K., Sarthi, J., Johnson, A.A., Toth, M.S., Chiu, K., Koduri, S., and Elefant, F. (2012). Tip60 HAT activity mediates APP induced lethality and apoptotic cell death in the CNS of a *Drosophila* Alzheimer's disease model. *PLoS ONE* 7, e41776.
- Polymenidou, M., Lagier-Tourenne, C., Hutt, K.R., Bennett, C.F., Cleveland, D.W., and Yeo, G.W. (2012). Misregulated RNA processing in amyotrophic lateral sclerosis. *Brain Res.* 1462, 3–15.
- Raj, B., Irimia, M., Braunschweig, U., Sterne-Weiler, T., O'Hanlon, D., Lin, Z.Y., Chen, G.I., Easton, L.E., Ule, J., Gingras, A.C., et al. (2014). A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol. Cell* 56, 90–103.
- Romero, P.R., Zaidi, S., Fang, Y.Y., Uversky, V.N., Radivojac, P., Oldfield, C.J., Cortese, M.S., Sickmeier, M., LeGall, T., Obradovic, Z., and Dunker, A.K. (2006). Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl. Acad. Sci. USA* 103, 8390–8395.
- Rusconi, F., Paganini, L., Braidà, D., Ponzone, L., Toffolo, E., Maroli, A., Landsberger, N., Bedogni, F., Turco, E., Pattini, L., et al. (2014). LSD1 neurospecific alternative splicing controls neuronal excitability in mouse models of epilepsy. *Cereb. Cortex*. Published online April 15, 2014. <http://dx.doi.org/10.1093/cercor/bhu070>.
- Sabo, S.L., Ikin, A.F., Buxbaum, J.D., and Greengard, P. (2003). The amyloid precursor protein and its regulatory protein, FE65, in growth cones and synapses in vitro and in vivo. *J. Neurosci.* 23, 5407–5415.
- Sofueva, S., Yaffe, E., Chan, W.C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* 32, 3119–3129.
- Stante, M., Minopoli, G., Passaro, F., Raia, M., Vecchio, L.D., and Russo, T. (2009). Fe65 is required for Tip60-directed histone H4 acetylation at DNA strand breaks. *Proc. Natl. Acad. Sci. USA* 106, 5093–5098.
- Tarpey, P.S., Stevens, C., Teague, J., Edkins, S., O'Meara, S., Avis, T., Barthorpe, S., Buck, G., Butler, A., Cole, J., et al. (2006). Mutations in the gene encoding the Sigma 2 subunit of the adaptor protein 1 complex, AP1S2, cause X-linked mental retardation. *Am. J. Hum. Genet.* 79, 1119–1124.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.
- Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., et al. (2005). Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet.* 37, 844–852.
- Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 474, 380–384.
- Volfovsky, N., Haas, B.J., and Salzberg, S.L. (2003). Computational discovery of internal micro-exons. *Genome Res.* 13 (6A), 1216–1221.
- Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718.
- Wang, B., Hu, Q., Hearn, M.G., Shimizu, K., Ware, C.B., Liggitt, D.H., Jin, L.W., Cool, B.H., Storm, D.R., and Martin, G.M. (2004). Isoform-specific knockout of FE65 leads to impaired learning and memory. *J. Neurosci. Res.* 75, 12–24.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476.
- Wang, Y., Zhang, M., Moon, C., Hu, Q., Wang, B., Martin, G., Sun, Z., and Wang, H. (2009). The APP-interacting protein FE65 is required for hippocampus-dependent learning and long-term potentiation. *Learn. Mem.* 16, 537–544.
- Weatheritt, R.J., Davey, N.E., and Gibson, T.J. (2012). Linear motifs confer functional diversity onto splice variants. *Nucleic Acids Res.* 40, 7123–7131.
- Wu, J., Anczuków, O., Krainer, A.R., Zhang, M.Q., and Zhang, C. (2013). OLEgo: fast and sensitive mapping of spliced mRNA-Seq reads using small seeds. *Nucleic Acids Res.* 41, 5149–5163.
- Wu, T.D., and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–1875.
- Zhang, Z., Pinto, A.M., Wan, L., Wang, W., Berg, M.G., Oliva, I., Singh, L.N., Dengler, C., Wei, Z., and Dreyfuss, G. (2013). Dysregulation of synaptogenesis genes antecedes motor neuron pathology in spinal muscular atrophy. *Proc. Natl. Acad. Sci. USA* 110, 19348–19353.
- Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keefe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* 34, 11929–11947.
- Zhu, L., Wang, X., Li, X.L., Towers, A., Cao, X., Wang, P., Bowman, R., Yang, H., Goldstein, J., Li, Y.J., and Jiang, Y.H. (2014). Epigenetic dysregulation of SHANK3 in brain tissues from individuals with autism spectrum disorders. *Hum. Mol. Genet.* 23, 1563–1578.

B Cell Super-Enhancers and Regulatory Clusters Recruit AID Tumorigenic Activity

Jason Qian,^{1,14} Qiao Wang,^{2,14} Marei Dose,^{1,14,*} Nathanael Pruett,^{1,14} Kyong-Rim Kieffer-Kwon,^{1,14} Wolfgang Resch,¹ Genqing Liang,¹ Zhonghui Tang,³ Ewy Mathé,¹ Christopher Benner,¹³ Wendy Dubois,⁵ Steevenson Nelson,¹ Laura Vian,¹ Thiago Y. Oliveira,² Mila Jankovic,² Ofir Hakim,⁶ Anna Gazumyan,² Rushad Pavri,⁷ Parirokh Awasthi,⁸ Bin Song,⁹ Geng Liu,⁹ Longyun Chen,⁹ Shida Zhu,⁹ Lionel Feigenbaum,⁸ Louis Staudt,¹⁰ Cornelis Murre,⁴ Yijun Ruan,³ Davide F. Robbiani,² Qiang Pan-Hammarström,^{11,15} Michel C. Nussenzweig,^{2,12,15} and Rafael Casellas^{1,5,15,*}

¹Genomics and Immunity, NIAMS, NIH, Bethesda, MD 20892, USA

²Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA

³Department of Genetic and Development Biology, Jackson Laboratory for Genomic Medicine, University of Connecticut, 400 Farmington, CT 06030, USA

⁴Division of Biological Sciences, Department of Molecular Biology, University of California, San Diego, La Jolla, CA 92093, USA

⁵Center of Cancer Research, NCI, NIH, Bethesda, MD 20892, USA

⁶Bar-Ilan University, Ramat-Gan 5290002, Israel

⁷Institute of Molecular Pathology (IMP), Vienna BioCenter, Doktor Bohr Gasse 7, Vienna 1030, Austria

⁸Science Applications International Corporation/Frederick, NCI-Frederick Cancer Research and Development Center, Frederick, MD 21702, USA

⁹Beijing Genomics Institute, Shenzhen, Shenzhen 518083, China

¹⁰Metabolism Branch, NCI, NIH, Bethesda, MD 20892, USA

¹¹Department of Laboratory Medicine, Karolinska Institutet, Karolinska University Hospital, Huddinge, 14186 Stockholm, Sweden

¹²HHMI, The Rockefeller University, New York, NY 10065, USA

¹³The Integrative Genomics and Bioinformatics Core, Salk Institute for Biological Studies, La Jolla, CA 92037, USA

¹⁴Co-first author

¹⁵Co-senior author

*Correspondence: marei.dose@nih.gov (M.D.), rafael.casellas@nih.gov (R.C.)

<http://dx.doi.org/10.1016/j.cell.2014.11.013>

SUMMARY

The antibody gene mutator activation-induced cytidine deaminase (AID) promiscuously damages oncogenes, leading to chromosomal translocations and tumorigenesis. Why nonimmunoglobulin loci are susceptible to AID activity is unknown. Here, we study AID-mediated lesions in the context of nuclear architecture and the B cell regulome. We show that AID targets are not randomly distributed across the genome but are predominantly grouped within super-enhancers and regulatory clusters. Unexpectedly, in these domains, AID deaminates active promoters and eRNA⁺ enhancers interconnected in some instances over megabases of linear chromatin. Using genome editing, we demonstrate that 3D-linked targets cooperate to recruit AID-mediated breaks. Furthermore, a comparison of hypermutation in mouse B cells, AID-induced kataegis in human lymphomas, and translocations in MEFs reveals that AID damages different genes in different cell types. Yet, in all cases, the targets are predominantly associated with topological complex, highly transcribed super-enhancers, demonstrating that these compartments are key mediators of AID recruitment.

INTRODUCTION

Although humans produce roughly equal numbers of B and T lymphocytes, up to 95% of lymphomas in the Western world are of B cell origin (Küppers, 2005). This overrepresentation originates in large part from misrepair of DNA lesions introduced by activation-induced cytidine deaminase (AID), a B cell-specific cytidine deaminase that initiates class switch recombination (CSR) and somatic hypermutation (SHM) of immunoglobulin (*Ig*) genes (Alt et al., 2013). Although AID preferentially targets *Ig* heavy and light chain loci, it also mutates and produces DNA breaks in non-*Ig* genes (Hakim et al., 2012; Liu et al., 2008; Robbiani et al., 2008). Among these off targets, a substantial number are oncogenes directly implicated in B cell lymphomagenesis, including *BCL6*, *Myc*, *MIR142*, *CD95*, *Pax5*, and *BCL7* (Chiarle et al., 2011; Hakim et al., 2012; Hasham et al., 2010; Kato et al., 2012; Klein et al., 2011; Müschen et al., 2000; Pasqualucci et al., 1998; Robbiani et al., 2009; Shen et al., 1998; Tsai et al., 2008). Recurrent DNA damage at these loci leads to oncogenic mutations and chromosomal translocations that activate proto-oncogenes by juxtaposing them to potent *Ig* enhancers (Nussenzweig and Nussenzweig, 2010). Accordingly, genetic ablation of AID markedly impairs the formation of *Ig*-translocations and the onset of B cell tumor development in mice (Kovalchuk et al., 2007, 2012; Ramiro et al., 2004; Robbiani et al., 2008; Takizawa et al., 2008).

Transcription facilitates AID targeting to *Ig* genes by at least three related mechanisms. First, *Ig* enhancers are required for

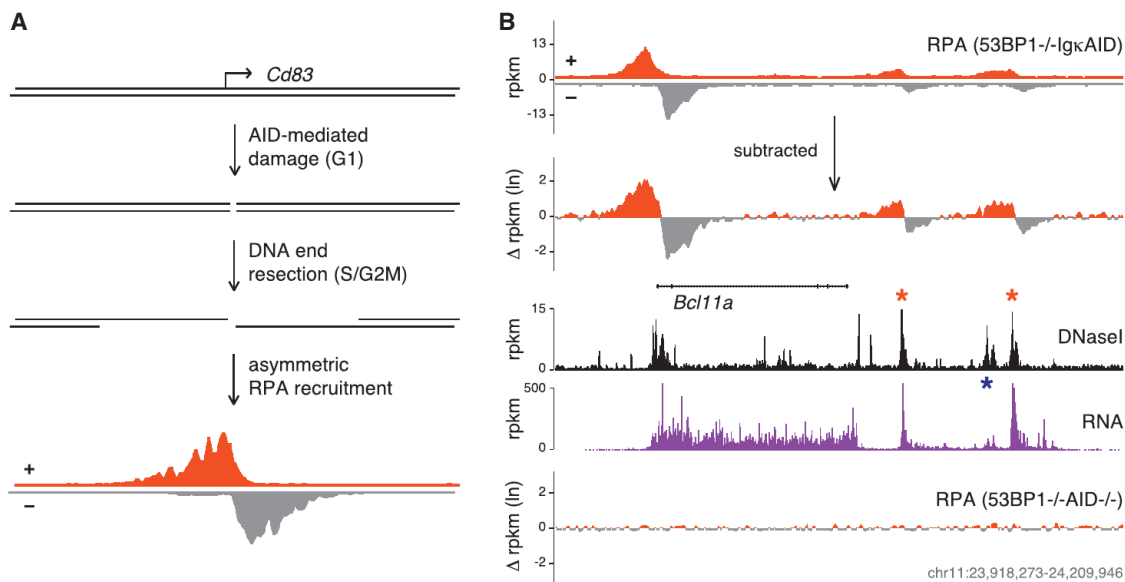


Figure 1. AID Damages Enhancer DNA

(A) Strategy to reveal AID-mediated breaks. In 53BP1^{-/-} cells DNA lesions at AID off-targets (e.g., *Cd83*) in G1 are resected in S and G2M by HR repair nucleases, leading to asymmetric RPA binding that can be detected by ChIP-Seq.

(B) The visualization of RPA-Seq was improved by plotting the difference in ChIP signals between + and - strands. An algorithm was developed to efficiently detect asymmetric RPA occupancy. The new approach reveals two additional AID targets at the *Bcl11a* locus that overlap with enhancer elements (highlighted with red asterisks). The nontargeted enhancer is marked with a blue asterisk. DNaseI, RNA (GRO-seq) (Chiarle et al., 2011), and RPA control (53BP1^{-/-}AID^{-/-}) tracks are provided.

See also Figure S1 and Table S1A.

hypermutation and recombination of both variable (V) domains and switch (S) DNA repeats that precede antibody gene constant (C) regions (Buerstedde et al., 2014). Second, transcription of S repeats leads to substantial RNA PolII pausing (Rajagopal et al., 2009; Wang et al., 2009), and Spt5, a PolII pausing factor, enables hypermutation and recombination by associating with AID (Pavri et al., 2010). Third, the RNA degrading exosome complex displaces nascent S transcripts thereby rendering both DNA strands accessible to deamination (Basu et al., 2011). Whether these or additional mechanisms are responsible for promiscuous AID activity at non-Ig loci is unknown.

Here, we examine promiscuous AID activity and its relationship to chromosome folding and the B cell regulome. We find that AID-mediated lesions occur predominantly within B cell super-enhancers and regulatory clusters. Furthermore, we show that the structural and transcriptional features of these domains help explain AID tumorigenic activity in the B cell compartment of mice and humans.

RESULTS

AID Damages Enhancer DNA

To study AID off-targeting activity, we made use of replication protein A chromatin immunoprecipitation (RPA-ChIP) that labels DNA breaks in the 53BP1^{-/-} background (Hakim et al., 2012). B cells isolated from these mice are defective for nonhomologous end joining (NHEJ), and AID-mediated lesions that are induced in G1 are aberrantly processed in S and G2M by homologous

recombination (Yamane et al., 2013). As a result, DNA-ends are resected leading to asymmetrical accumulation of RPA and Rad51 around DNA breaks and these proteins can be detected by chromatin immunoprecipitation (Figure 1A)

To improve the sensitivity of the assay, we developed an algorithm that detects asymmetric RPA recruitment with high precision, and the difference in ChIP signals between upper (+) and lower (-) DNA strands was plotted on a log scale (Figure 1B). The new approach revealed 92 additional genomic sites associated with RPA in 53BP1^{-/-}IgκAID B cells (236 total targets; Table S1A available online). Conversely, we detected a single RPA asymmetric peak in 53BP1^{-/-}AID^{-/-} cells (not shown). At the *Bcl11a* locus, for instance, we found two additional sites downstream of the promoter (120 and 180 kb away) that display asymmetric RPA accumulation in the presence of AID but not in its absence (Figure 1B). Notably, a fraction of the peaks (33, or 14%) did not overlap with TSSs but were associated with DNaseI hypersensitive sites corresponding to B cell enhancers (red asterisks in Figure 1B) (Kieffer-Kwon et al., 2013). Consistent with this interpretation, AID targets distal from TSSs displayed the epigenetic signature of active enhancers: H2AZ^{low}H3K4me3^{low}H3K4me1^{high} (Kouzine et al., 2013; not shown). Thus, in addition to promoter proximal sequences, AID damages enhancer DNA.

Nuclear Compartmentalization of AID Activity

AID activity is confined to the interphase nucleus (Petersen et al., 2001), where the genome is partitioned into a hierarchy of

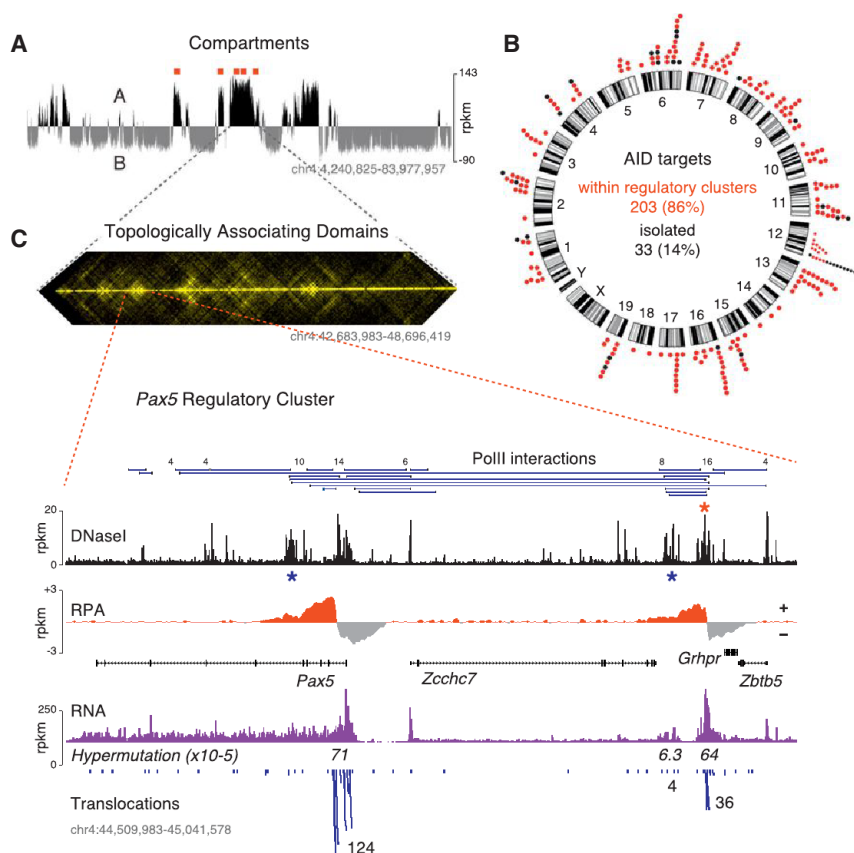


Figure 2. Tethering and Compartmentalization of AID Targets in the Mouse Genome

(A) AID targets are largely found within A compartments (black upper track) as defined by Hi-C. Red dots identify the location of damaged loci within the genomic domain. The Hi-C data was obtained from pro-B cells. All other experiments involving mouse B cells in the manuscript were done with activated B cells.

(B) Circos plot shows the genome-wide distribution of AID targets that are either tethered within regulatory clusters (red dots) or isolated (black dots).

(C) Upper: heat map of *cis*-interaction frequencies revealing TADs within the domain chr4:42,683,983-48,696,419. Lower: *Pax5* gene regulatory cluster, as defined by PolII long-range interactions. The targeted promoter is associated with nondamaged (blue asterisks) and damaged (red asterisk) enhancers. DNaseI hypersensitivity, RNA, hypermutation, and chromosomal translocations (TC-Seq) are also shown. The number of interactions is provided above the ChIA-PET links. See also Figure S1.

structures, including A-B compartments, topologically associating domains (TADs), and clusters of interactive gene regulatory elements (Gibcus and Dekker, 2013). The finding that both promoters and enhancers undergo AID-mediated damage suggests that AID targets might also be clustered in the B cell nucleus. In support of this idea, nearly half of all targets (110 of 236) were located within ~90 kb of each other, a distance that is markedly different from a random model (~4 Mb, Figure S1A). Prompted by these observations, we analyzed the distribution of RPA+ sites in the context of genome folding, as defined by chromosome conformation capture (3C) techniques.

Hi-C maps from pro-B cells (Lin et al., 2012) revealed that 96% of AID targets (233 of 236) are located within A compartments (Table S1A; Figure 2A). These compartments are generally gene-rich, DNaseI-hypersensitive, and transcriptionally active (Lieberman-Aiden et al., 2009), features that agree well with AID's preference for transcribed chromatin.

In eukaryotes, TADs divide A-B compartments into nuclear subdomains containing clusters of multiple regulatory elements tethered by long-range interactions (Gibcus and Dekker, 2013; Li et al., 2012). To examine the distribution of AID targets vis-à-vis this architecture we made use of a PolII ChIA-PET map from activated B cells (Kieffer-Kwon et al., 2013). This technique combines PolII ChIP with 3C technology to define the promoter-enhancer interactome. Remarkably, while 47% of active promoters in B lymphocytes are not anchored in regulatory clusters, (Table S1A), 86% of AID targets were preferentially tethered to

neighboring promoters and enhancers within regulatory clusters ($p < 10^{-15}$, Figure 2B and Experimental Procedures). In some cases, these clusters connected multiple AID targets. For instance, at the *Pax5* locus the targeted promoter was linked by long-range interactions with three enhancer domains, one of which (~250 kb away) was also damaged by AID (Figure 2C). Likewise, the targeted *Ly6a*, *Ly6e*, and *Rohema* promoters in chromosome 15 formed a topological cluster spanning ~100 kb (Figure S1B). Importantly, the vast majority of AID targets (84%) were tethered to regulatory elements within the same TADs (e.g., *Pax5* cluster, Figure 2C), consistent with the notion that these domains restrict chromatin mobility (Gibcus and Dekker, 2013). A notable exception was the histone H1 gene family, where AID targets from two noncontiguous compartments physically associated over 2.1 Mb (Figure S1C). We conclude that AID preferentially damages promoters and enhancers tethered by long-range interactions within gene regulatory clusters.

AID Targeting Is Largely Confined to B Cell Super-Enhancers

Super-enhancers (SEs) or stretch enhancers were recently identified as a special subset of regulatory elements (Hnisz et al., 2013; Lovén et al., 2013; Parker et al., 2013; Whyte et al., 2013). They represent exceptionally large enhancer domains primarily associated with highly transcribed genes controlling cell identity. Because of the known correlation between transcription and AID activity, we asked whether regulatory clusters targeted by AID might represent SE domains. To this end, we used H3K27Ac and a published algorithm (Whyte et al., 2013) to catalog SEs in stimulated B cells. Consistent with the high degree of activation in the presence of LPS+IL-4, we uncovered 1,003 SEs in cultured B cells (Figure S2A). By comparison, 13% of 86

human tissues surveyed displayed >1,000 SEs (Hnisz et al., 2013). In agreement with such studies, activated B cell SEs spanned DNA regions an order of magnitude greater than conventional enhancers, and they were densely occupied by the Mediator complex (Figure S2B).

At all three *Ig* loci, AID-mediated damage occurred within SE domains interconnected by long-range interactions (Figures 3A and S2C). Remarkably, 76% (179 of 236) of all AID targets were linked to SEs, a significant enrichment over what is expected by chance ($p < 1 \times 10^{-15}$, see Experimental Procedures). As an example, both the *Aicda*- and *Apobec1*-targeted genes are interconnected within the same SE (Figure 3C). Thus, AID on- and off-targeting activity occurs primarily within SE domains.

A key characteristic of SEs is that they are largely cell-type specific. Consistent with this, more than 50% of AID-targeted SEs were only present in stimulated B cells when compared to 18 primary mouse cells and tissues (Figure S2D). The analysis included SEs from developing pro-B cells (Whyte et al., 2013), which only displayed 32% overlap with activated counterparts (Figure S2D). Hence, most AID-mediated damage occurs within SEs acquired during development.

Approximately 80% (824 of 1,003) of B cell SEs did not harbor AID-mediated damage (Figure 3B). Notably, SEs containing AID targets could be distinguished from nontargeted ones in that they were more accessible (higher H3K27Ac, $p = 1 \times 10^{-25}$, Figure 3D), larger in size ($p = 3 \times 10^{-9}$, Figure 3E), and their associated promoters were transcribed at higher levels ($p = 4 \times 10^{-10}$, Figure 3F). In addition, the extent of 3D connectivity was significantly higher at targeted SEs ($p = 3 \times 10^{-17}$, Figure 3G). We conclude that AID targets are preferentially associated with SEs displaying a high degree of accessibility, transcription, and structural complexity.

Functional Attributes of AID-Targeted Regulatory Elements

Within SEs, genes undergoing AID-mediated damage are linked to both targeted and nontargeted elements. For instance, of 11 enhancers associated with *Myc*, only two showed asymmetric RPA occupancy (Figure 4A). To characterize features that might distinguish these two enhancer groups, we measured hypersensitivity to DNaseI but found no significant differences ($p = 0.9$, Figure 4B). Conversely, targeted enhancers were consistently transcribed, as determined by GRO-Seq analysis ($p = 3 \times 10^{-5}$, Figure 4C). For instance, of the two enhancers upstream of *Pax5*, only the one displaying high levels of eRNA synthesis was associated with RPA, chromosomal translocations, and somatic hypermutation (Figure 2C). Additional examples at the *Bcl11a* locus are provided in Figure 1B. Similarly, the RPA+ *Myc* enhancers at the mid-point of *Pvt1* were transcribed at higher levels compared to those lacking RPA (Figure 4A). Of note, *Igk* translocations involving this particular *Myc* enhancer cluster are selected during plasmacytomagenesis (Huppi et al., 2011).

Consistent with eRNA synthesis, PolII and PolIII long-range interactions were significantly higher at enhancers associated with AID-mediated lesions ($p = 2 \times 10^{-6}$, Figure 4D and not shown). The PolII stalling factor Spt5, implicated in AID recruitment (Pavri et al., 2010), was also enriched in RPA+ enhancers ($p = 6 \times 10^{-4}$,

Figure 4E). Importantly, these features were particularly prominent at hypermutated *Igh* E μ and *Igk* E α enhancers, whereas they were consistently low at the nontargeted *Igl* E3-1 and E3-1 s enhancers (Figures 4C–4E; Table S1B). Conversely, no differences were found in the recruitment of CTCF, a factor involved in nuclear architecture ($p = 0.03$, Figure 4F). A separate analysis showed that these same features distinguished AID-targeted from nontargeted promoters (Figure S3A). Thus, AID preferentially deaminates transcriptionally active promoters and enhancers that engage in frequent long-range interactions.

Interacting Targets within SEs Cooperate to Recruit AID Activity

The clustering of AID targets in the mouse genome suggests that they may cooperate or synergize to recruit AID to SE domains. To directly test this idea we asked whether a nontargeted, but otherwise highly transcribed promoter could recruit hypermutation when linked to a damaged gene cluster. To this end, we inserted the ubiquitin-C (*Ubc*) gene promoter from chromosome 5 in lieu of the *Il4ra* promoter in chromosome 7 to generate *Il4ra*^{u/u} mice (Figure S3B). In activated B cells, *Il4ra* and flanking *Nsmce1* and *Il21r* overlap with SEs and interact extensively creating a multiple-promoter gene cluster (Figure 5A). In the presence of AID, all three genes undergo DNA double-strand breaks (Figure 5A), whereas no damage is detected at *Ubc* (Figure S3C).

Fluorocytometric analysis of *Il4ra*^{u/u} and *Il4ra*^{+/+} B cells showed comparable levels of cell surface *Il4ra* receptor (Figure S3D). Consistent with this result, knockin B cells proliferated normally and underwent wild-type levels of γ 1 recombination (Figures S3E and S3F). Importantly, H3K27Ac and RNA-Seq showed little or no differences in SE location or expression of *Nsmce1*, *Il4ra*, or *Il21r* between the two cell types (Figures 5B, 5C, and S4A). To measure chromatin contacts at the knockin allele we applied an improved version of 4C-Seq that characterizes local architecture at high resolution (van de Werken et al., 2012). The analysis showed that the knocked-in *Ubc* promoter associates with flanking *Nsmce1* and *Il21r* genes at wild-type frequencies (Figure S4B). Similar results were obtained when using the *Il21r* promoter as bait (Figure S4C). Thus, neither transcription nor the architecture of the *Nsmce1*-*Il4ra*-*Il21r* locus appeared disrupted following promoter replacement.

To directly assess AID activity we bred the *Il4ra*^u allele into the *Ung*^{-/-}*Igk*AID background, which enables measurement of hypermutation in *ex-vivo* cultures (Hakim et al., 2012). *Il4ra*^u/*Ung*^{-/-}*Igk*AID and *Il4ra*^{+/+}*Ung*^{-/-}*Igk*AID B cells were stimulated for 7 days and mutations downstream of *Ubc* were assessed at chromosomes 5 (native configuration) and 7 (knockin alleles). Consistent with the lack of DNA breaks at *Ubc* in chromosome 5 (Figure S3C), biological triplicates revealed background mutation at this site, comparable to the average PCR error rate measured in AID^{-/-} cells (SHM(f) = 13.6×10^{-5} versus 8.7×10^{-5} ; Figure 5D; Table S1B). Notably, in *Il4ra*^{u/u} cells *Ubc* displayed a significant increase in mutation frequency in chromosome 7 compared to its native site (SHM(f) = 59.2×10^{-5} , fold change = 4.3, $p = 0.0005$, Figure 5D). This mutation frequency was nearly that of *Il4ra* in wild-type cells (80.5×10^{-5} , Figure 5D). *Mir142*, *Pim1*, and *Myc*, which are not directly associated with the *Il4ra* locus, showed no significant changes in

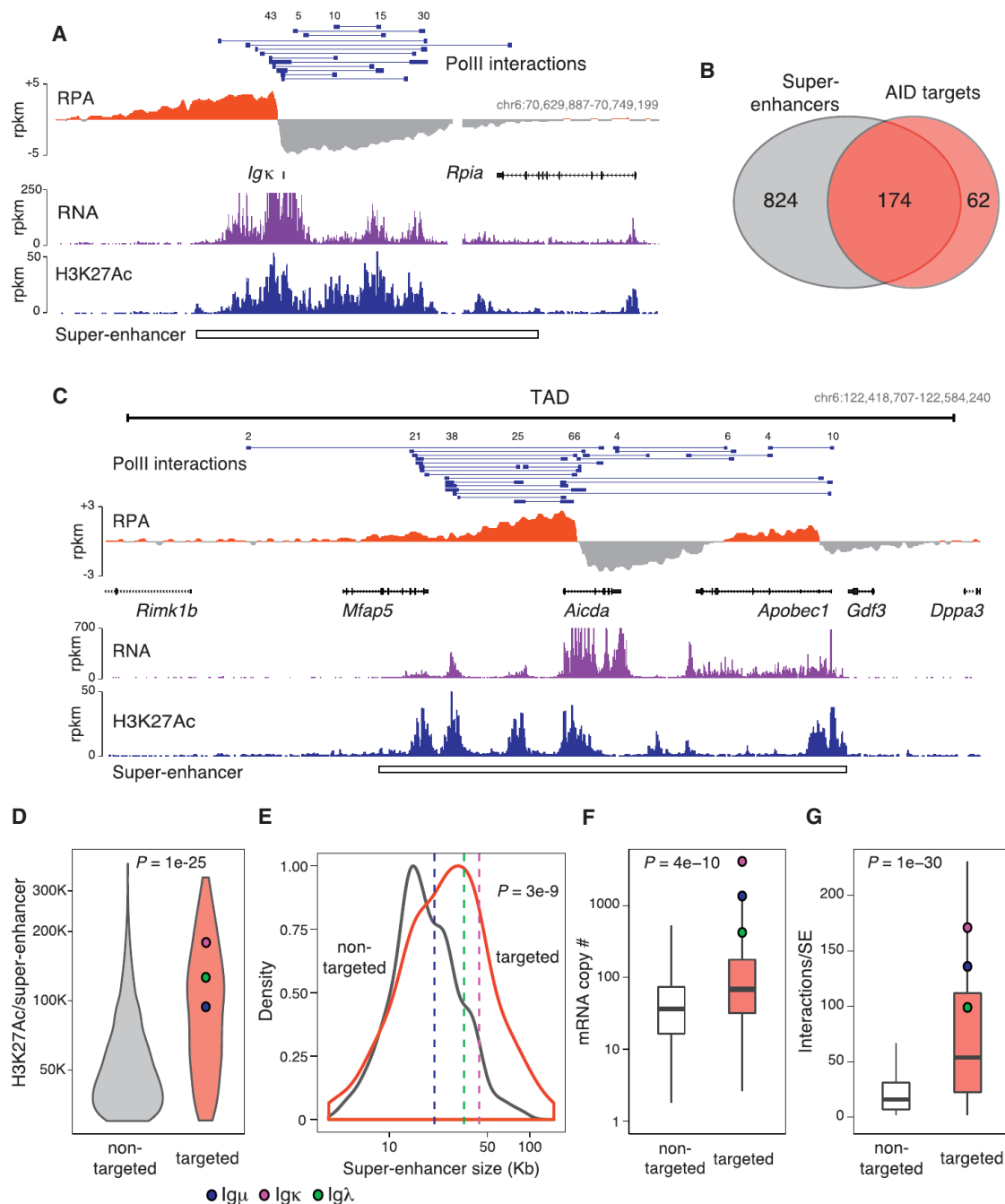


Figure 3. AID-Targeted Regulatory Clusters Are Predominantly Associated with B Cell SEs

(A) AID activity at the *Igκ* locus occurs within a 65 kb SE domain displaying long-range chromatin interactions. PolII interactions, RPA, RNA, and H3K27Ac profiles are provided.

(B) Venn diagram showing the fraction of AID targets associated with B cell SEs.

(C) Example of AID off-targeted SEs at the *Aicda-Apobec1* TAD in chromosome 6.

(D) H3K27Ac signal at targeted and nontargeted SEs. *Igμ* (blue, chr12:114640978-114669901), *Igκ* (magenta, chr6:70659188-70724456), and *Igλ* (green, chr.16:19002804-19067747) SEs are highlighted.

(E) Size distribution of total constituent enhancers in targeted (red line) or nontargeted (black line) SEs.

(F and G) Box plots showing the absolute expression or PolII-mediated connections at targeted (red) and nontargeted (open) SEs. Data are represented as the mean \pm SEM.

See also Figure S2.

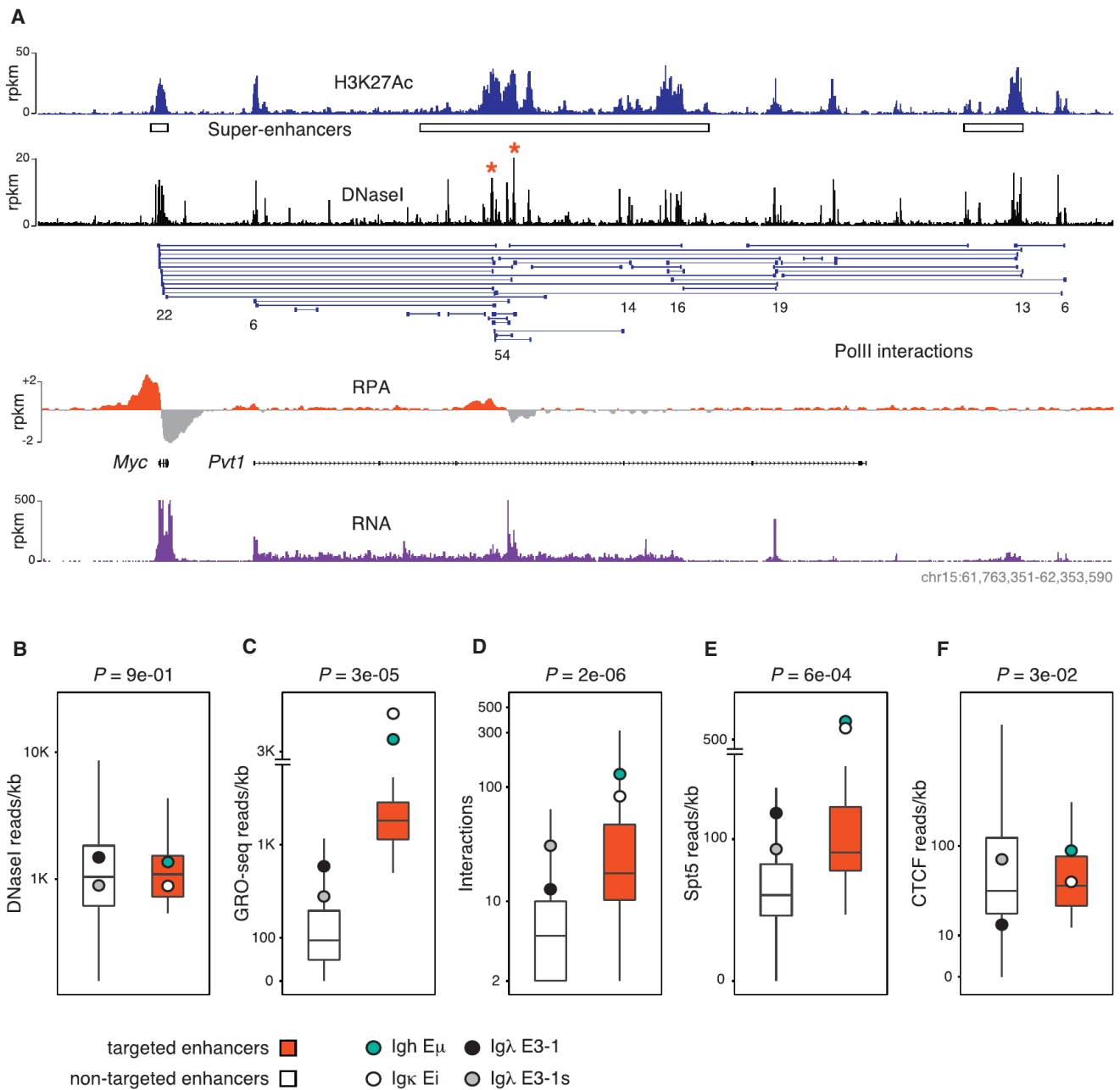


Figure 4. Defining Features of Targeted Enhancers

(A) *Myc* locus showing the distribution of SEs (H3K27Ac-Seq), enhancers (DNaseI-Seq), PolII long-range interactions (ChIA-PET), AID-mediated damage (RPA-Seq), and RNA synthesis (GRO-Seq). AID-targeted enhancers are denoted with red asterisks.

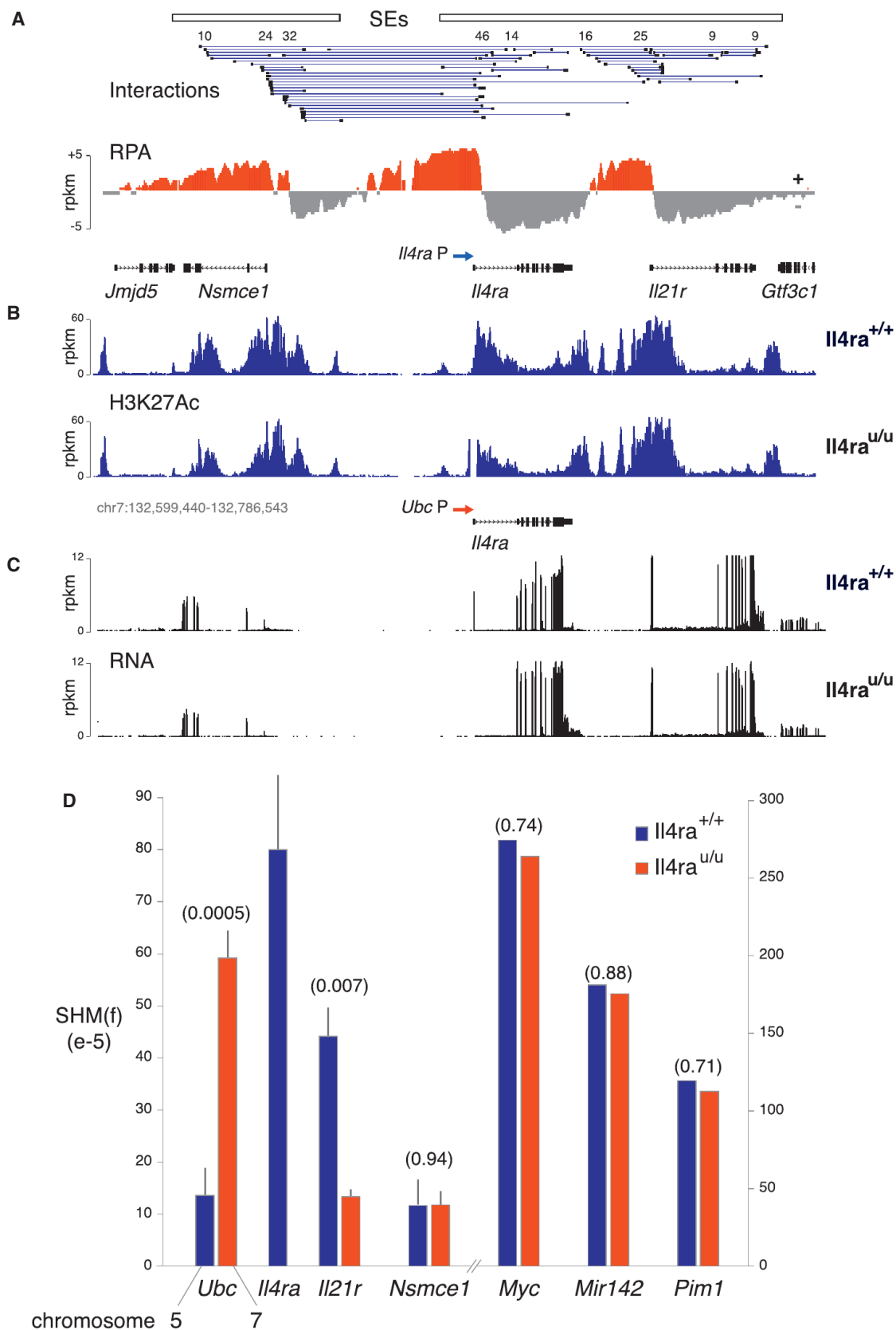
(B–F) Box plots comparing the extent of protein recruitment (B), DNaseI-Seq, eRNA synthesis (C), GRO-Seq, PolII interactions (D), PETs, Spt5 (E), and CTCF occupancy (F) at targeted (red boxes) and nontargeted (open boxes) enhancers. Data are represented as the mean \pm SEM.

See also Figure S3.

hypermutation following gene targeting (fold change = 1.0–1.1, $p > 0.7$; Figure 5D; Table S1B). Hence, regulatory sequences at the *Nsmce1-Il4ra-Il21r* locus promote AID activity at *Ubc*.

To determine whether the *Il4ra* promoter also facilitates AID activity at flanking genes, we measured hypermutation at *Il21r* and *Nsmce1* in wild-type and knockin B cells. At *Nsmce1*, muta-

tion could not be detected above background (Figure 5D). Conversely, at *Il21r*, where SHM(f) was 44.3×10^{-5} in wild-type cells, we observed a statistically significant decrease in *Il4ra*^{u/u} (13.2×10^{-5} , $p = 0.007$, Figure 5D), indicating that replacement of the *Il4ra* promoter for *Ubc* has a negative effect on mutation of *Il21r* more than 50 kb downstream. We conclude



(legend on next page)

that both the *Il4ra* promoter and additional regulatory sequences at the *Nsmce1-Il4ra-Il21r* gene cluster enable off-targeting hypermutation by AID. The findings thus support a model where topologically linked elements within targeted SEs cooperate to recruit AID-mediated damage.

AID Targets in Human Lymphomas Overlap with Regulatory Clusters and SEs

Despite the known link between AID activity and human B cell tumor development (Klein and Dalla-Favera, 2008; Seifert et al., 2013), a comprehensive map of AID targets in the human genome is lacking. To directly address this question and to validate our findings in mouse B cells we mapped AID activity in the Ramos Burkitt's lymphoma line and in primary diffuse large B cell lymphoma (DLBCL). These tumors derive from germinal center or postgerminal center B cells and frequently display evidence of AID activity (Lossos et al., 2004; Pasqualucci et al., 2004; Sale and Neuberger, 1998). To efficiently detect hypermutation in Ramos we developed a deep-sequencing assay (SHM-Seq) by disrupting the mismatch repair gene *MSH2* by genome editing with a cassette expressing AID and Ugi, an inhibitor of the base excision repair factor Ung (Figure S5A). The resulting cell line is therefore both Ung- and Msh2-deficient, a combination that in mouse B cells leads to high levels of AID-mediated transition mutations at *Ig* and off-target loci (Hakim et al., 2012; Liu et al., 2008). Following 300 days of culture, the targeted cell line was single-cell sorted, individual clones were expanded, and DNA associated with H3K4me3, a histone mark that overlaps with AID activity (Yamane et al., 2011), was isolated and microsequenced (Figure S5A). Nontargeted and AID^{-/-} Ramos cells were used as controls.

Analysis of 26 clones revealed 11,344 mutations relative to nontargeted and AID^{-/-} controls. As expected, 92% of the substitutions were transitions. At *IGH* we detected 1,474 mutations (SHM(f) = 1.0×10^{-2}), mostly downstream of VDJ and S μ promoters (Figure 6A; Table S1C). Likewise, the *IGH*-translocated *MYC* allele was highly mutated (SHM(f) = 5.0×10^{-3} , Figure 6B). The nontranslocated *MYC* allele was also targeted but at a frequency ~20-fold lower (SHM(f) = 2.2×10^{-4} , not shown). Other oncogenes often targeted in human lymphomas showed evidence of AID activity, including *MIR142*, *BCL6*, *BCL7A*, *MSH6*, and *ID3* (Table S1C). In total, 60 sites were hypermutated with high confidence, including four conventional enhancers (false discovery rate [FDR] < 10^{-16} , see Experimental Procedures).

Our mouse studies were performed with B cells overexpressing AID and in ex-vivo cultures, where SHM is limited. To map AID activity in unmanipulated cells we next performed whole-

genome sequencing (~40× coverage) of ten DLBCL primary tumors isolated from lymph node biopsies. Somatic substitutions were defined by sequencing normal blood cells from the same patients. A total of 145,997 mutations were identified concomitant with deletions, insertions, amplifications, and chromosomal translocations. To classify AID hypermutation targets with high confidence we took advantage of the processive nature of AID deamination, which can generate clusters of transition mutations in individual clones (Lada et al., 2012; Taylor et al., 2013). These mutation showers or kataegis were recently uncovered by whole-genome sequencing of B cell and nonhematopoietic tumors (Alexandrov et al., 2013; Bolli et al., 2014; Chen et al., 2014; Nik-Zainal et al., 2012; Sakofsky et al., 2014). In the latter, particularly in breast tumors, kataegis was ascribed to processive deamination by the AID-related enzyme APOBEC3B (Alexandrov et al., 2013; Taylor et al., 2013).

We identified 105 kataegic sites in DLBCL associated with 30 genes (Table S1D). Four features implicated AID in the etiology of these mutation clusters. First, 82% of kataegis overlapped with transcribed promoter sequences, AID's preferred targeting domain (Figure S5B). This is in stark contrast to published non-B cell tumors (Alexandrov et al., 2013), where <6% of the kataegis were associated with TSSs ($p < 1 \times 10^{-10}$, Figure S5B). Second, also in contrast to other tumors, kataegis in DLBCL were recurrent, in that they always involved the *IG* loci and in most instances other mouse AID targets such as *PIM1*, *PAX5*, *RHOH*, *CIITA*, *MIR142*, *BCL6*, and the AID gene itself *AICDA* (Figure 6C; Table S1D). Third, 71% of the mutations were C > T transitions, consistent with the notion that kataegis results from DNA replication over cytidine deamination of resected DNA (Sakofsky et al., 2014; Taylor et al., 2013). Fourth, targeted cytidines bear the hallmark of AID activity (Taylor et al., 2013), i.e., they occur in a sequence context that recapitulates AID's preference for WRCY hotspots (Chi-square test $p < 1 \times 10^{-15}$, Figure 6D) (Rogozin and Kolchanov, 1992). Conversely, mutated Cs in breast tumors only differed from the genome average in that they were preceded mostly by a T (Figure 6D), consistent with the deamination profile of APOBEC3B (Alexandrov et al., 2013; Burns et al., 2013). These results thus support the proposal that kataegis in human lymphomas stem from AID activity.

We next characterized AID targets from Burkitt's and DLBCL tumors in the context of nuclear architecture and SEs. To this end we mapped PolII ChIA-PET and H3K27Ac in Ramos and used germinal center B cells isolated from human tonsils as substitutes for primary DLBCL (see Experimental Procedures). Consistent with the mouse results we found a strong overlap between hypermutated genes and SE domains (57%–70%,

Figure 5. Tethered Regulatory Elements Cooperate to Recruit AID Activity

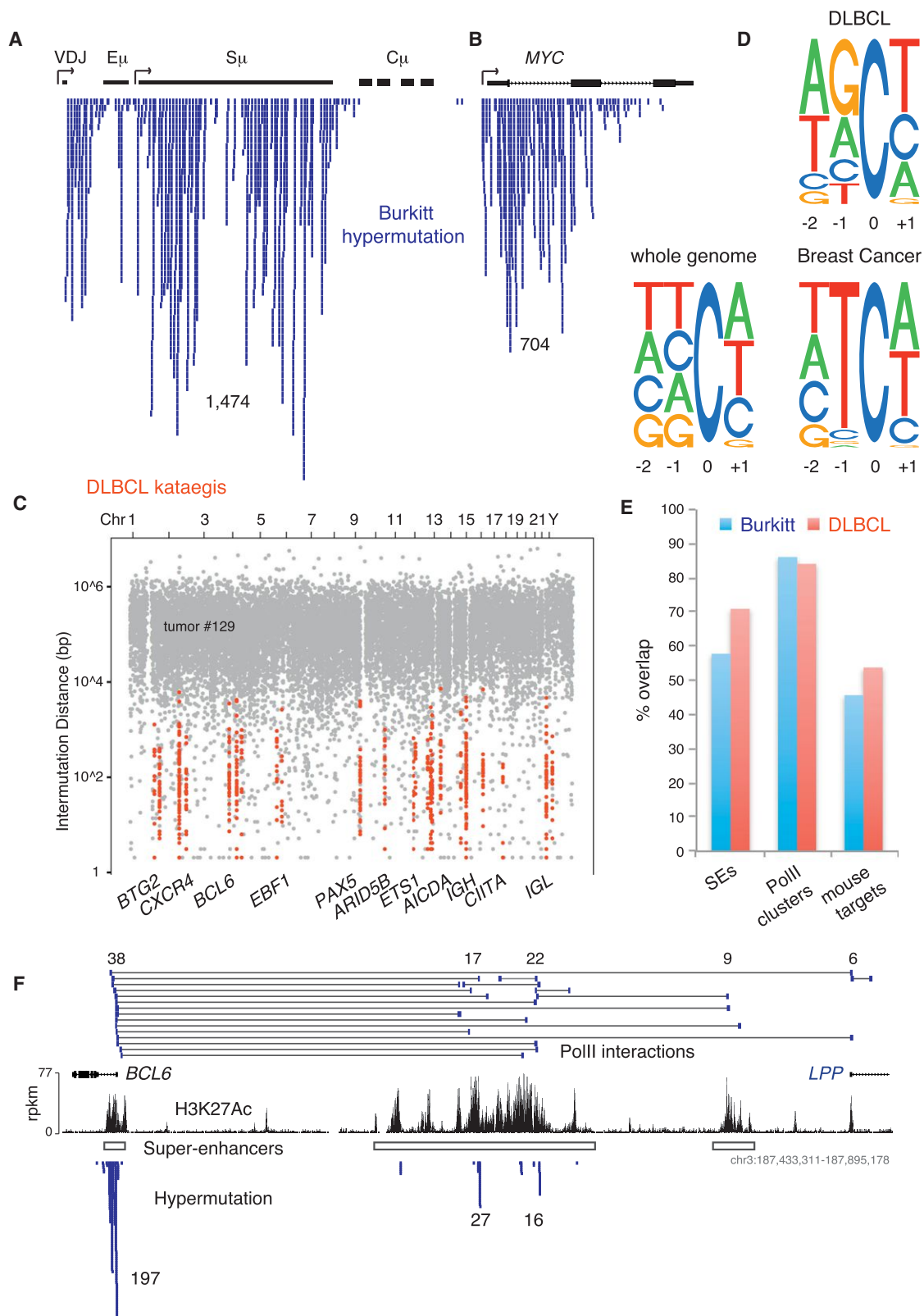
(A) *Il4ra*, *Il21r*, and *Nsmce1* form a promoter-gene cluster on mouse chromosome 7. Long-range interactions, DNA damage, and SEs are shown. *Il4ra*^{u/u} knockin mice were created by replacing the *Il4ra* promoter (P, blue arrow) for that of *Ubc* (red arrow).

(B) H3K27Ac in wild-type and knockin mouse B cells.

(C) mRNA expression (plotted as rpkm values).

(D) Hypermutation frequency at *Ubc*, *Il4ra*, *Il21r*, *Nsmce1*, *Myc*, *Mir142*, and *Pim1* genes was measured in *Il4ra*^{+/+} (blue bars) and *Il4ra*^{u/u} (red bars) activated B cells. P values shown were calculated with Student's t test for triplicates experiments (*Ubc*, *Il21r*, *Nsmce1*) and Fisher's exact test (*Myc*, *Mir142*, *Pim1*) for single experiments. Hypermutation at *Ubc* was measured on chromosome 5 in *Il4ra*^{+/+} (blue bar) and on chromosome 7 in *Il4ra*^{u/u} (red bar). Data are represented as the mean \pm SEM.

See also Figures S3 and S4 and Table S1B.



(legend on next page)

Figure 6E, $p < 1 \times 10^{-15}$, see Experimental Procedures). Furthermore, 83%–85% of the targets were anchored by PolII long-range interactions (Figure 6E). For instance, at the *BCL7A* gene regulatory cluster in Ramos both the promoter and upstream enhancers were hypermutated (Figure S5C). Another notable example was the *BCL6* promoter and a linked SE domain >250 kb upstream (Figure 6F). Importantly, while only the *BCL6* promoter was associated with kataegis in DLBCL, a survey of 26 genomes from other primary human lymphomas (Alexandrov et al., 2013) revealed the presence of kataegis at the upstream SE domain (Figure S5D). Altogether, the results demonstrate that both in mouse and human B cells AID mutates tethered regulatory elements associated with SEs and regulatory clusters.

AID Targets a Specific Microenvironment Rather Than a Determined Set of Genes

The kataegis and SHM-Seq analyses of B cell tumors revealed that 57%–85% of human AID targets overlap with SEs and regulatory clusters, whereas the overlap with mouse targets was only 45%–53% (Figure 6E). The strong inference is that rather than mutating a specific set of genes, AID targets topologically complex, highly transcribed domains. To directly test this idea we mapped AID-induced translocations in MEFs using TC-Seq (Klein et al., 2011). Primary AID^{-/-} MEFs carrying I-SceI sites at *Myc* and *Igh* (*Myc⁺Igh⁺*AID^{-/-}) were transduced with I-SceI alone or I-SceI and AID. A total of 15,272 unique, mappable rearrangements to *Myc⁺* were captured from 40 million AID^{-/-} MEFs, and 28,265 from 40 million AID-expressing MEFs (2 libraries each, Table S1E). Similar to B cells (Klein et al., 2011), a large fraction (20%–44%) of the rearrangements in MEFs occurred in *cis* within a 250 kb window around I-SceI (Figure S6A). Furthermore, translocations were associated with genes more frequently than predicted by a random model (binomial test $p < 0.0001$, Table S1E). Using stringent criteria, we identified 29 and 43 AID-dependent translocation hotspots in MEFs and B cells, respectively (Table S1F). Remarkably, while the majority of these hotspots were genic (>84%), only three (11%) were shared between fibroblasts and lymphocytes (Figure 7A). This result indicates that the cell type alters the landscape of genomic rearrangements induced by AID.

Because the spatial organization of the genome is not random but compartmentalized (Lieberman-Aiden et al., 2009), it is possible that the cell type-restricted translocation to *Myc⁺* results from differences in nuclear organization. However, 4C-Seq showed that the *Myc* interactome in fibroblasts and lymphocytes was highly similar (Pearson's $\rho = 0.88$, Figure S6B), consistent with the observation that nuclear interactions do not correlate

with the frequency of AID-mediated translocations (Hakim et al., 2012).

To explore the contribution of transcription to cell type-specific targeting, we next measured RNA synthesis. We found that, in general, genes associated with translocation hotspots displayed higher transcription in the respective cell type (Figure S6C). For example, *Pax5* and *Cd83* were only targeted and expressed in B cells, while MEF-specific hotspots *Ctcf* and *Wisp1* were only transcribed in fibroblasts (Figure S6D). Furthermore, while *Myc* was frequently translocated to the *Igh* I-SceI site in MEFs, we failed to detect rearrangements to S domains, which in fibroblasts are transcriptionally silent (Figure S6E). To assess whether differential AID targeting was also associated with SE domains we analyzed publicly available H3K27Ac profiles. We found that, similar to results obtained with B cells, AID activity at hotspot genes in MEFs occurred largely within the context of SEs (71%, $p < 1 \times 10^{-10}$, Figure 7B and Experimental Procedures). Importantly, this correlation applied to genes that were expressed in both cell types but were targeted in only one of them, such as *Flnb* on chromosome 14 (Figure 7C) and *Pim1* on chromosome 17 (Figure S7A). Altogether, the findings demonstrate that whereas AID damages a different set of genes in MEFs and B cells, in both cell types the targets are preferentially associated with SEs domains.

DISCUSSION

Recurrent translocation to non-*Ig* loci in B cell cancers is due in part to DNA damage by AID (Chiarle et al., 2011; Hakim et al., 2012; Klein et al., 2011; Zhang et al., 2012). However, the genomic features responsible for recruiting DNA damage are unknown. Our studies of mouse B cells, human lymphomas, and MEFs reveal that a major unifying property of AID targets is that they are predominantly clustered within highly active SEs and regulatory clusters (Figure S7B). As discussed below, the functional and architectural properties of these domains help explain why their associated genes are susceptible to AID tumorigenic activity.

SEs represent a special subset of regulatory clusters, where chromatin accessibility and transcriptional activity are an order of magnitude higher than at other active sites (Parker et al., 2013; Whyte et al., 2013). Both accessibility and transcription have long been recognized as prerequisites to *Ig* gene deamination (Alt et al., 2013). Our experiments show that along with size and long-range interconnectivity, the presence of a SE can differentiate targeted from nontargeted regulatory elements. For instance, a model based on these combined features can

Figure 6. AID Targets in Human Lymphomas Are Associated with Long-Range Chromatin Interactions and SEs

(A and B) The SHM-Seq protocol detects AID-mediated hypermutation in Ramos B cells, including at the *IGH* (A) and *MYC* (B) loci.

(C) Rainfall plot displaying the distance between neighboring mutations across the genome of a DLBCL primary tumor (#129). Kataegic domains of clustered mutations are depicted with red dots. Some of the genes associated with kataegis are highlighted.

(D) Representation of sequence context at positions -2, -1, and +1 flanking mutated Cs in DLBCL or breast cancer kataegis. The average context of Cs in the entire human genome is also shown.

(E) Percent overlap between hypermutated genes from Ramos Burkitt's lymphoma (blue bars) or primary DLBCL (red bars) in SEs (left), PolII long-range interactions (middle), or mouse AID targets (right).

(F) AID hypermutation of the *BL6* regulatory cluster in Ramos cells. SEs, PolII long-range interactions, and hypermutation are provided.

See also Figure S5 and Tables S1C and S1D.

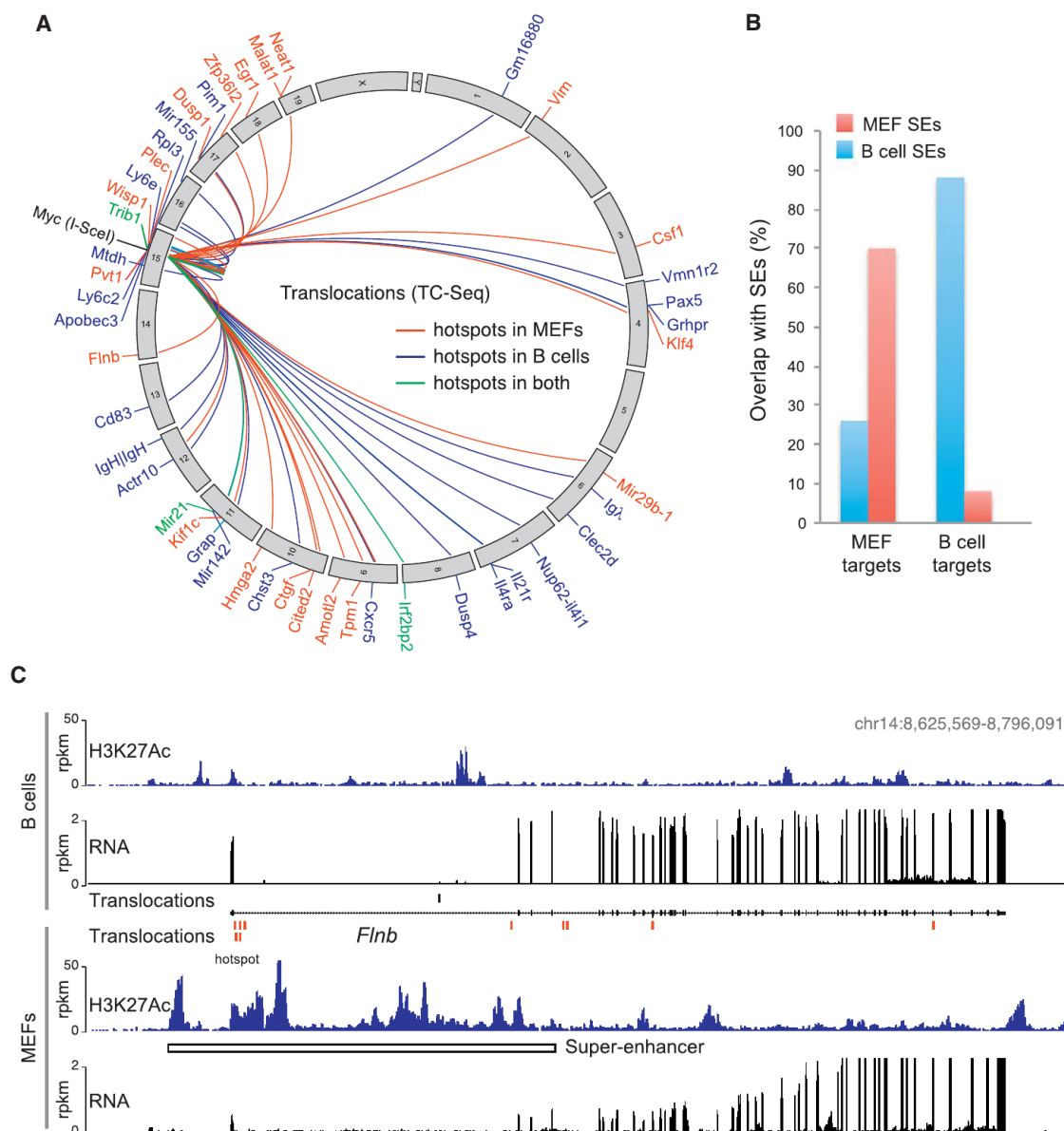


Figure 7. AID Damages Different Genes in Different Cell Types

(A) Circos diagram showing hotspots of AID-dependent chromosome translocations to *Myc*^{I-SceI} in MEFs and B cells. Hotspots only present in B cells (blue lines), MEFs (red lines), or both cell types (green lines) are provided.

(B) Overlap of AID targets in MEFs (red bars) or B cells (blue bars) with SEs.

(C) *Myc* translocations to *Flnb* are primarily detected in MEFs (red bars), where the gene is associated with a SE domain. Conversely, a single translocation is detected in B cells (black bar).

See also Figures S6 and S7 and Tables S1E and S1F.

accurately predict 91% of mouse AID targets at a false discovery rate of 9% (Figure S7C; Experimental Procedures). The underlying assumption is that, as a group, these properties help create a nuclear microenvironment highly suitable to AID-mediated deamination. The fact that our data cannot predict AID targeting in its totality implies that additional parameters might also be at play. Specific transcription factors for instance have been shown to facilitate AID recruitment to *Ig* genes (Buerstedde et al., 2014).

Small RNA processing by the Exosome complex is another example (Pefanis et al., 2014). Furthermore, in the accompanying paper, Alt and colleagues uncovers a strong correlation between convergent transcription and AID-mediated damage (Meng et al., 2014 in this issue of *Cell*).

Another unexpected finding is that within targeted SEs AID not only damages promoter proximal sequences but also cognate enhancers. These are invariably transcribed and more frequently

anchored by PolII long-range interactions. Both features likely render enhancer DNA accessible to cytidine deamination and double-strand break formation.

The link between AID activity and SEs sheds new light on the class of genes damaged in activated and germinal center B cells. Genome-wide maps of SHM, DNA breaks, and chromosomal translocations have consistently uncovered two sets of genes enriched among AID targets: oncogenes involved in proliferation and apoptosis (e.g., *Myc*, *Pim1*, *Jund*, *Bcl2*) and genes that feature prominently in B cell development and activation (*Pax5*, *Cd79b*, *Aicda*, *Irf8*, *Bach2*, *Nfkb*). Although AID's predilection for these gene groups has been unclear, they fit well with the observation that in all tissues examined so far, SEs largely control expression of cell identity genes as well as oncogenes that regulate cell cycle and differentiation. Examples of these are pluripotency genes in ES cells, genes critical for islet function in the pancreas, and *MYC* in multiple myeloma (Lovén et al., 2013; Parker et al., 2013; Whyte et al., 2013). By the same token, our TC-Seq analysis showed that targeted SEs in MEFs control expression of genes critical for fibroblast proliferation and maturation (e.g., *Ctgf*, *Wisp1*, *Amotl2*).

Another defining feature of SEs is that their constituent regulatory elements work in cooperation or synergistically to drive gene expression (Lovén et al., 2013). Our knockin experiments between the nontargeted *Ubc* promoter and the *Nsmce1-Il4ra-Il21r* targeted gene cluster provide compelling evidence that cooperation is also key to promiscuous AID-mediated damage. This feature helps explain why AID targets are clustered in the B cell genome. At the same time, it suggests that only networks of functionally cooperating elements can create the proper conditions for AID promiscuous activity. It is important to point out that these conditions are not exclusive to SE domains, but that they also typify highly interactive regulatory clusters not directly associated with SEs (e.g., H1 gene family). The *Ubc-Il4ra* experiment also provides a rationale to earlier observations showing that heterologous promoters not typically damaged in germinal centers (e.g., β -globin, B29, or PolII promoters) can recruit hypermutation when juxtaposed to *Ig* enhancers (Betz et al., 1994; Fukita et al., 1998; Tumas-Brundage and Manser, 1997). In both cases, AID exploits long-range interactions to act at a distance on nontargeted sequences.

In conclusion, rather than targeting a predetermined gene set, AID tumorigenic activity is focused on nuclear microenvironments that share a common set of architectural, transcriptional, and regulatory features.

EXPERIMENTAL PROCEDURES

Extended Experimental Procedures are provided in the Supplemental Information section.

4C-Seq

The 4C assay was performed as previously described van de Werken et al. (2012) with minor modifications. Ten million mouse B cells were crosslinked in 2% formaldehyde at 37°C for 10 min. The reaction was quenched by the addition of glycine (final concentration of 0.125 M). Cells were then washed with cold PBS and lysed (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.2% NP-40, 1× complete protease inhibitors [Roche]) at 4°C for 1 hr. Nuclei were incubated at 65°C for 30 min, 37°C for 30 min in 500 μ l of restriction buffer (New England

BioLabs DpnII buffer) containing 0.3% SDS. To sequester SDS, Triton X-100 was then added to a final concentration of 1.8%. DNA digestion was performed with 400 U of DpnII (New England Biolabs) at 37°C overnight. After heat inactivation (65°C for 30 min), the reaction was diluted to a final volume of 7 ml with ligation buffer containing 100 U T4 DNA Ligase (Roche) and incubated at 16°C overnight. Samples were then treated with 500 μ g Proteinase K (Ambion) and incubated overnight at 65°C to reverse formaldehyde crosslinking. DNA was then purified by phenol extraction and ethanol precipitation. For circularization, the ligation junctions were digested with Csp6I (Fermentas) at 37°C overnight. After enzyme inactivation and phenol extraction, the DNA was religated in a 7 ml volume (1,000 U T4 DNA Ligase, Roche). Three micrograms of 4C library DNA was amplified with Expand Long template PCR System (Roche). Thermal cycle conditions were DNA denaturing for 2 min at 94°C, followed by 30 cycles of 15 s at 94°C, 1 min at 58°C, 3 min at 68°C, and a final step of 7 min at 68°C. Baits were amplified with inverse PCR primers as follows: *Il4ra* with DpnII: $\text{4C } 5\text{'-TCAGGTAGTTCATGGGATC-3'}$, *Il4ra_Csp6I* $5\text{'-ATCTCTGCACCAGACATCAG-3'}$ and *Il21r* with IL21r_DpnII $\text{CCAGACCTACTTAGCAGATC}$, and IL21r_Csp6I: $\text{ACTTAGACACTGCTCAGCTG}$. 4C-amplified DNA was microsequenced with the Illumina platform.

ACCESSION NUMBERS

The two Gene Expression Omnibus (GEO) and one Sequence Read Archive (SRA) accession numbers for the deep-sequencing data reported in this paper are GSE62063, GSE61523, and SRP046243, respectively.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.013>.

AUTHOR CONTRIBUTIONS

Q.P.-H., M.C.N., and R.C. designed experiments. J.Q., Q.W., N.P., and K.-R.K.-K. performed most key experiments. S.N., L.V., M.J., O.H., A.G., D.F.R., and R.P. performed experiments. G. Liang., P.A., W.D., and L.F. generated and maintained *Il4ra* mice. M.D., W.R., E.M., Z.T., T.Y.O., and C.B. performed bioinformatics. B.S., G. Liu., L.C., and S.Z. analyzed WGSs. L.S., C.M., and Y.R. provided expertise advice. M.C.N. and R.C. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Daniel Hodson for isolating GC B cells, Ethan Tyler for designing Figure S7B, Gustavo Gutierrez for technical assistance with deep-sequencing, David Bosque and Tom Eisenreich for managing the mouse colonies, and Jim Simone, Klara Velinzon, and Yelena Shatalina for FACS sorting. This work was supported by the Intramural Research Programs of the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) and the National Cancer Institute (NCI), the Jackson Laboratory fund (JAX19020120), the European Research Council (ERC) (242551-ImmunoSwitch), and NIH grant AI037526 to M.C.N. M.C.N. is a Howard Hughes Medical Institute (HHMI) Investigator. Animal experiments were performed according to NIH and Rockefeller approved protocols. This study used the high-performance computational capabilities of NIH Helix Systems (<http://helix.nih.gov>).

Received: May 26, 2014

Revised: September 26, 2014

Accepted: October 30, 2014

Published: December 4, 2014

REFERENCES

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al.; Australian

- Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.
- Alt, F.W., Zhang, Y., Meng, F.L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell* 152, 417–429.
- Basu, U., Meng, F.L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R., et al. (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* 144, 353–363.
- Betz, A.G., Milstein, C., González-Fernández, A., Pannell, R., Larson, T., and Neuberger, M.S. (1994). Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. *Cell* 77, 239–248.
- Bolli, N., Avet-Loiseau, H., Wedge, D.C., Van Loo, P., Alexandrov, L.B., Martincorena, I., Dawson, K.J., Iorio, F., Nik-Zainal, S., Bignell, G.R., et al. (2014). Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat. Commun.* 5, 2997.
- Buerstedde, J.M., Alinikula, J., Arakawa, H., McDonald, J.J., and Schatz, D.G. (2014). Targeting of somatic hypermutation by immunoglobulin enhancer and enhancer-like sequences. *PLoS Biol.* 12, e1001831.
- Burns, M.B., Lackey, L., Carpenter, M.A., Rathore, A., Land, A.M., Leonard, B., Refsland, E.W., Kotandeniya, D., Tretyakova, N., Nikas, J.B., et al. (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* 494, 366–370.
- Chen, X., Bahrami, A., Pappo, A., Easton, J., Dalton, J., Hedlund, E., Ellison, D., Shurtleff, S., Wu, G., Wei, L., et al.; St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project (2014). Recurrent somatic structural variations contribute to tumorigenesis in pediatric osteosarcoma. *Cell Reports* 7, 104–112.
- Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molinie, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., et al. (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* 147, 107–119.
- Fukita, Y., Jacobs, H., and Rajewsky, K. (1998). Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity* 9, 105–114.
- Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. *Mol. Cell* 49, 773–782.
- Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K.-R., Jankovic, M., Oliveira, T., Bothmer, A., Voss, T.C., Ansarah-Sobrinho, C., et al. (2012). DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature* 484, 69–74.
- Hasham, M.G., Donghia, N.M., Coffey, E., Maynard, J., Snow, K.J., Ames, J., Wilpan, R.Y., He, Y., King, B.L., and Mills, K.D. (2010). Widespread genomic breaks generated by activation-induced cytidine deaminase are prevented by homologous recombination. *Nat. Immunol.* 11, 820–826.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–947.
- Huppi, K., Pitt, J., Wahlberg, B., and Caplen, N.J. (2011). Genomic instability and mouse microRNAs. *Toxicol. Mech. Methods* 21, 325–333.
- Kato, L., Begum, N.A., Burroughs, A.M., Doi, T., Kawai, J., Daub, C.O., Kawaguchi, T., Matsuda, F., Hayashizaki, Y., and Honjo, T. (2012). Nonimmunoglobulin target loci of activation-induced cytidine deaminase (AID) share unique features with immunoglobulin genes. *Proc. Natl. Acad. Sci. USA* 109, 2479–2484.
- Kieffer-Kwon, K.-R., Tang, Z., Mathe, E., Qian, J., Sung, M.H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* 155, 1507–1520.
- Klein, U., and Dalla-Favera, R. (2008). Germinal centres: role in B-cell physiology and malignancy. *Nat. Rev. Immunol.* 8, 22–33.
- Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* 147, 95–106.
- Kouzine, F., Wojtowicz, D., Yamane, A., Resch, W., Kieffer-Kwon, K.-R., Bandle, R., Nelson, S., Nakahashi, H., Awasthi, P., Feigenbaum, L., et al. (2013). Global regulation of promoter melting in naive lymphocytes. *Cell* 153, 988–999.
- Kovalchuk, A.L., duBois, W., Mushinski, E., McNeil, N.E., Hirt, C., Qi, C.F., Li, Z., Janz, S., Honjo, T., Muramatsu, M., et al. (2007). AID-deficient Bcl-xL transgenic mice develop delayed atypical plasma cell tumors with unusual Ig/Myc chromosomal rearrangements. *J. Exp. Med.* 204, 2989–3001.
- Kovalchuk, A.L., Ansarah-Sobrinho, C., Hakim, O., Resch, W., Tolarová, H., Dubois, W., Yamane, A., Takizawa, M., Klein, I., Hager, G.L., et al. (2012). Mouse model of endemic Burkitt translocations reveals the long-range boundaries of Ig-mediated oncogene deregulation. *Proc. Natl. Acad. Sci. USA* 109, 10972–10977.
- Küppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. *Nat. Rev. Cancer* 5, 251–262.
- Lada, A.G., Dhar, A., Boissy, R.J., Hirano, M., Rubel, A.A., Rogozin, I.B., and Pavlov, Y.I. (2012). AID/APOBEC cytosine deaminase induces genome-wide kataegis. *Biol. Direct* 7, 47.
- Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84–98.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
- Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C. (2012). Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat. Immunol.* 13, 1196–1204.
- Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. *Nature* 451, 841–845.
- Lossos, I.S., Levy, R., and Alizadeh, A.A. (2004). AID is expressed in germinal center B-cell-like and activated B-cell-like diffuse large-cell lymphomas and is not correlated with intracolon heterogeneity. *Leukemia* 18, 1775–1779.
- Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153, 320–334.
- Meng, F.-L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.-R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberger, D., et al. (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159. Published online December 4, 2014. <http://dx.doi.org/10.1016/j.cell.2014.11.014>.
- Müschen, M., Re, D., Jungnickel, B., Diehl, V., Rajewsky, K., and Küppers, R. (2000). Somatic mutation of the CD95 gene in human B cells as a side-effect of the germinal center reaction. *J. Exp. Med.* 192, 1833–1840.
- Nik-Zainal, S., Alexandrov, L.B., Wedge, D.C., Van Loo, P., Greenman, C.D., Raine, K., Jones, D., Hinton, J., Marshall, J., Stebbings, L.A., et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell* 149, 979–993.
- Nussenzweig, A., and Nussenzweig, M.C. (2010). Origin of chromosomal translocations in lymphoid cancer. *Cell* 141, 27–38.
- Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., et al.; NISC Comparative Sequencing Program; National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors; NISC Comparative Sequencing Program Authors (2013). Chromatin stretch

enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. USA* **110**, 17921–17926.

Pasqualucci, L., Migliazza, A., Fracchiolla, N., William, C., Neri, A., Baldini, L., Chaganti, R.S., Klein, U., Küppers, R., Rajewsky, K., and Dalla-Favera, R. (1998). BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. *Proc. Natl. Acad. Sci. USA* **95**, 11816–11821.

Pasqualucci, L., Guglielmino, R., Houldsworth, J., Mohr, J., Aoufouchi, S., Polakiewicz, R., Chaganti, R.S., and Dalla-Favera, R. (2004). Expression of the AID protein in normal and neoplastic B cells. *Blood* **104**, 3318–3325.

Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., et al. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* **143**, 122–133.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* **514**, 389–393.

Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H.T., Difilippantonio, M.J., Wilson, P.C., Hanitsch, L., Celeste, A., Muramatsu, M., Pilch, D.R., et al. (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature* **414**, 660–665.

Rajagopal, D., Maul, R.W., Ghosh, A., Chakraborty, T., Khamlichi, A.A., Sen, R., and Gearhart, P.J. (2009). Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. *J. Exp. Med.* **206**, 1237–1244.

Ramiro, A.R., Jankovic, M., Eisenreich, T., Difilippantonio, S., Chen-Kiang, S., Muramatsu, M., Honjo, T., Nussenzweig, A., and Nussenzweig, M.C. (2004). AID is required for c-myc/IgH chromosome translocations in vivo. *Cell* **118**, 431–438.

Robbiani, D.F., Bothmer, A., Callen, E., Reina San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A., et al. (2008). AID is required for the chromosomal translocations in c-myc that lead to c-myc/IgH translocations. *Cell* **135**, 1028–1038.

Robbiani, D.F., Bunting, S., Feldhahn, N., Bothmer, A., Camps, J., Deroubaix, S., McBride, K.M., Klein, I.A., Stone, G., Eisenreich, T.R., et al. (2009). AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Mol. Cell* **36**, 631–641.

Rogozin, I.B., and Kolchanov, N.A. (1992). Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim. Biophys. Acta* **1171**, 11–18.

Sakofsky, C.J., Roberts, S.A., Malc, E., Mieczkowski, P.A., Resnick, M.A., Gordenin, D.A., and Malkova, A. (2014). Break-induced replication is a source of mutation clusters underlying kataegis. *Cell Reports* **7**, 1640–1648.

Sale, J.E., and Neuberger, M.S. (1998). TdT-accessible breaks are scattered over the immunoglobulin V domain in a constitutively hypermutating B cell line. *Immunity* **9**, 859–869.

Seifert, M., Scholtysik, R., and Küppers, R. (2013). Origin and pathogenesis of B cell lymphomas. *Methods Mol. Biol.* **971**, 1–25.

Shen, H.M., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science* **280**, 1750–1752.

Takizawa, M., Tolarová, H., Li, Z., Dubois, W., Lim, S., Callen, E., Franco, S., Mosaico, M., Feigenbaum, L., Alt, F.W., et al. (2008). AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. *J. Exp. Med.* **205**, 1949–1957.

Taylor, B.J., Nik-Zainal, S., Wu, Y.L., Stebbings, L.A., Raine, K., Campbell, P.J., Rada, C., Stratton, M.R., and Neuberger, M.S. (2013). DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *eLife* **2**, e00534.

Tsai, A.G., Lu, H., Raghavan, S.C., Muschen, M., Hsieh, C.L., and Lieber, M.R. (2008). Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. *Cell* **135**, 1130–1142.

Tumas-Brundage, K., and Manser, T. (1997). The transcriptional promoter regulates hypermutation of the antibody heavy chain locus. *J. Exp. Med.* **185**, 239–250.

van de Werken, H.J., Landan, G., Holwerda, S.J., Hoichman, M., Klous, P., Chachik, R., Splinter, E., Valdes-Quezada, C., Oz, Y., Bouwman, B.A., et al. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969–972.

Wang, L., Wuerffel, R., Feldman, S., Khamlichi, A.A., and Kenter, A.L. (2009). S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J. Exp. Med.* **206**, 1817–1830.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319.

Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat. Immunol.* **12**, 62–69.

Yamane, A., Robbiani, D.F., Resch, W., Bothmer, A., Nakahashi, H., Oliveira, T., Rommel, P.C., Brown, E.J., Nussenzweig, A., Nussenzweig, M.C., and Casellas, R. (2013). RPA accumulation during class switch recombination represents 5′-3′ DNA-end resection during the S-G2/M phase of the cell cycle. *Cell Reports* **3**, 138–147.

Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908–921.

Convergent Transcription at Intragenic Super-Enhancers Targets AID-Initiated Genomic Instability

Fei-Long Meng,^{1,7} Zhou Du,^{1,2,7} Alexander Federation,^{3,7} Jiazhi Hu,¹ Qiao Wang,⁴ Kyong-Rim Kieffer-Kwon,⁵ Robin M. Meyers,¹ Corina Amor,¹ Caitlyn R. Wasserman,¹ Donna Neuberger,⁶ Rafael Casellas,⁵ Michel C. Nussenzweig,⁴ James E. Bradner,^{3,8,*} X. Shirley Liu,^{6,8} and Frederick W. Alt^{1,8,*}

¹Howard Hughes Medical Institute, Program in Cellular and Molecular Medicine, Boston Children's Hospital, and Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

²Department of Bioinformatics, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China

³Department of Medical Oncology, Dana-Farber Cancer Institute, and Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

⁴Howard Hughes Medical Institute, Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA

⁵Genomics and Immunity, NIAMS, and Center of Cancer Research, NCI, National Institutes of Health, Bethesda, MD 20892, USA

⁶Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, MA 02115, USA

⁷Co-first author

⁸Co-senior author

*Correspondence: james_bradner@dfci.harvard.edu (J.E.B.), alt@enders.tch.harvard.edu (F.W.A.)

<http://dx.doi.org/10.1016/j.cell.2014.11.014>

SUMMARY

Activation-induced cytidine deaminase (AID) initiates both somatic hypermutation (SHM) for antibody affinity maturation and DNA breakage for antibody class switch recombination (CSR) via transcription-dependent cytidine deamination of single-stranded DNA targets. Though largely specific for immunoglobulin genes, AID also acts on a limited set of off-targets, generating oncogenic translocations and mutations that contribute to B cell lymphoma. How AID is recruited to off-targets has been a long-standing mystery. Based on deep GRO-seq studies of mouse and human B lineage cells activated for CSR or SHM, we report that most robust AID off-target translocations occur within highly focal regions of target genes in which sense and antisense transcription converge. Moreover, we found that such AID-targeting “convergent” transcription arises from antisense transcription that emanates from super-enhancers within sense transcribed gene bodies. Our findings provide an explanation for AID off-targeting to a small subset of mostly lineage-specific genes in activated B cells.

INTRODUCTION

The B cell antigen receptor (BCR) is comprised of immunoglobulin (Ig) heavy (IgH) and light (IgL) chains. In response to antigen activation, B lymphocytes in peripheral lymphoid organs undergo somatic hypermutation (SHM) and IgH class switch recombination (CSR) and ultimately secrete their BCR as an anti-

body. SHM diversifies antibody repertoires by introducing high-frequency mutations into IgH and IgL variable region exons (Di Noia and Neuberger, 2007). SHM occurs in germinal centers (GCs) of peripheral lymphoid tissues, where B cells are selected for mutations that generate BCRs with increased antigen affinity (Victoria and Nussenzweig, 2012). IgH CSR involves generation and joining of *IgH* locus DSBs in switch (S) regions that precede various sets of *IgH* C_H exons (C_Hs) to replace the initially expressed C_H with a downstream C_H, thereby producing antibodies with different effector functions (Matthews et al., 2014). Both SHM and CSR are initiated by activation-induced cytidine deaminase (AID) (Muramatsu et al., 2000), which deaminates cytosine to uridine on single-stranded DNA (ssDNA) (Di Noia and Neuberger, 2007). Mismatches created by these deaminated cytidines are processed into mutations or DSBs during SHM and CSR, respectively, through a process that employs activities of normal base excision or mismatch repair pathways (Di Noia and Neuberger, 2007).

Within target sequences, AID cytidine deamination focuses on 3–4 bp “SHM” motifs that are greatly enriched in S regions and in portions of variable region exons that encode antigen-binding sites (Di Noia and Neuberger, 2007). Transcription is required for AID targeting during SHM and CSR (Alt et al., 2013; Storb, 2014). In this regard, SHM of V(D)J exons in GC B cells begins ~150 bp downstream of the transcription start site (TSS) and tapers off 1–2 kb downstream (Liu and Schatz, 2009). Likewise, each C_H has a promoter upstream of the S region that upon induction by external signals generates transcription through the S region and, thereby, targets AID (Matthews et al., 2014). Mouse and human S regions also have a highly G-rich nontemplate strand that upon transcription forms stable R-loops that provide ssDNA to augment AID targeting (Matthews et al., 2014; Alt et al., 2013). RNA polymerase II (Pol II) has been implicated in directing AID to Ig gene SHM and CSR targets through a transcription

coupled mechanism (Storb, 2014) that involves AID association with the Spt5 transcription cofactor in the context of Pol II stalling (Pavri et al., 2010). R loops or other aspects of repetitive S region structure may augment AID access by promoting Pol II stalling (Rajagopal et al., 2009; Wang et al., 2009). Once AID is recruited to Ig targets, replication protein A (RPA) and the RNA exosome RNA degradation complex contribute to generating requisite ssDNA substrates (Basu et al., 2011; Matthews et al., 2014; Panfili et al., 2014).

Beyond Ig gene targets, AID initiates recurrent mutations or DSBs in a small subset of non-Ig genes collectively termed AID “off-target” genes (Pasqualucci et al., 2001; Chiarle et al., 2011; Klein et al., 2011; Liu et al., 2008). Off-target AID activity promotes translocations between Ig loci and cellular oncogenes, as well as SHMs of oncogenes associated with B cell lymphomas (Alt et al., 2013; Küppers and Dalla-Favera, 2001). Identification of AID off-targets has been facilitated by genome-wide translocation cloning methods (Chiarle et al., 2011; Klein et al., 2011) and other large-scale approaches (Liu et al., 2008; Yamane et al., 2011). In general, AID activity occurs at much lower levels on off-targets than on Ig genes (Liu and Schatz, 2009; Yamane et al., 2011; Chiarle et al., 2011; Klein et al., 2011), likely due to specialized AID-targeting features of the latter. AID off-target sequences are not enriched in AID hotspot motifs relative to the genome in general (Duke et al., 2013). Consistent with a role for transcription, AID off-target activity is most abundant on transcribed genes downstream of their TSSs (Pasqualucci et al., 2001; Liu et al., 2008; Chiarle et al., 2011; Klein et al., 2011). However, transcription per se is not sufficient to target AID, as most transcribed genes are not AID off-targets (Alt et al., 2013; Liu and Schatz, 2009). Next-generation sequencing studies revealed unexpected transcriptional features, including divergent sense and antisense transcription at TSSs (Wu and Sharp, 2013; Adelman and Lis, 2012) and frequent promoter proximal Pol II pausing (Adelman and Lis, 2012). But, divergent transcription (DivT) from TSSs occurs in over half of all genes and generally does not map directly to sites of AID off-target activity (Chiarle et al., 2011; see below). Likewise, transcriptional pausing alone cannot explain AID off-targeting, because more than 30% of transcribed genes have paused Pol II (Adelman and Lis, 2012). Thus, mechanisms that lead to recurrent AID targeting may arise from previously unrecognized transcriptional or epigenetic determinants (Alt et al., 2013).

Global run-on sequencing (GRO-seq) detects nascent transcripts generated by transcriptionally engaged RNA polymerases (Core et al., 2008). GRO-seq revealed that a large fraction of intergenic regions are transcribed, with a subset emanating from transcriptional enhancers (Wang et al., 2011). Enhancers are sequence-defined, *cis*-regulatory elements that influence target gene expression irrespective of orientation (Levine et al., 2014). Both enhancers within genes (intragenic) and intergenic enhancers may regulate target promoters locally and over long distances (Levine et al., 2014). Active enhancer sequences are commonly transcribed by RNA Pol II generating so-called “enhancer RNAs” (eRNAs), and transcription arising from enhancers is often divergent, with both sense and antisense transcription emanating from enhancer elements (Natoli and Andrau, 2012; Wang et al., 2011). Various regulatory functions have been

ascribed to eRNAs and other noncoding RNAs (Lam et al., 2014), however, much of noncoding RNA biology is not fully understood.

Enhancers are comprised of discrete or clustered transcription factor binding sequences. A common feature of active enhancers is chromatin that is characteristically modified by acetylation (e.g., histone 3 lysine 27; H3K27Ac) and methylation (e.g., histone 3 lysine 4 mono-methylation; H3K4me1) (Creyghton et al., 2010). An unexpected asymmetry in the regional allocation of enhancer factors and enrichment for enhancer marks within and unique to each mammalian cell type studied revealed a subset of so-called super-enhancers (SEs) that feature clusters of hyperacetylated and actively transcribed enhancers that, on average, are 10-fold longer than other “typical” enhancers (Whyte et al., 2013; Lovén et al., 2013). Like locus control regions, SEs regulate genes involved in specialized cellular function (Parker et al., 2013) and are found within or adjacent to lineage-specifying transcription factor genes (Whyte et al., 2013; Hnisz et al., 2013). In cancer, SEs frequently enforce oncogene expression (Lovén et al., 2013) and, thereby, contribute to tumor pathogenesis. For example, translocations that juxtapose *c-myc* to the IgH 3′ regulatory region, a known SE (Delmore et al., 2011; Chapuy et al., 2013), promote B cell lymphoma by activating *c-myc* over long distances (Gostissa et al., 2009). In this context, selectively blocking SE activity with bromodomain and extra-terminal domain (BET) inhibitors is a promising cancer therapeutic strategy (Delmore et al., 2011; Lovén et al., 2013; Chapuy et al., 2013).

Here, we report that the majority of detectable AID off-target activity in a variety of mouse and human lymphoid or nonlymphoid cell types occurs within focal regions of overlapping sense/antisense transcription within intragenic SEs.

RESULTS

Deep GRO-Seq Transcription Profiles of Naive, GC, and CSR-Activated B Cells

To elucidate transcriptional features that influence AID targeting genome-wide, we applied GRO-seq to splenic naive, GC, and CSR-activated B cells at much greater depth than done previously. Naive splenic B cells were purified (Figure S1A available online) and then cultured in the presence of α CD40 plus interleukin-4 (IL4) for 60 hr to stimulate AID induction and CSR to IgG1 and IgE (Figure S1A). Splenic GC B cells were purified from sheep red blood cell immunized mice (Figure S1A) and confirmed to be >90% pure (Figures S1B–S1D). Three independent GRO-seq biological replicates were performed for each cell type and gave highly reproducible results (Figure S1E). Transcription profiles of over 20,000 genes revealed distinct (but overlapping) gene expression patterns for each cell type that were further classified by gene ontology terms (Figure S1G; Table S1). As expected (Core et al., 2008; Chiarle et al., 2011), GRO-seq revealed divergent sense and antisense transcription at TSSs of over 50% of the genes in each of the three cell types (Figures 1 and S1F). In-depth examination of sense transcription profiles of several “signature” genes illustrates the specificity of purified cell populations. For example, *Aicda* sense transcription reflects AID protein expression in the three cell types, with high levels in GC B cells and activated B cells; but none detectable

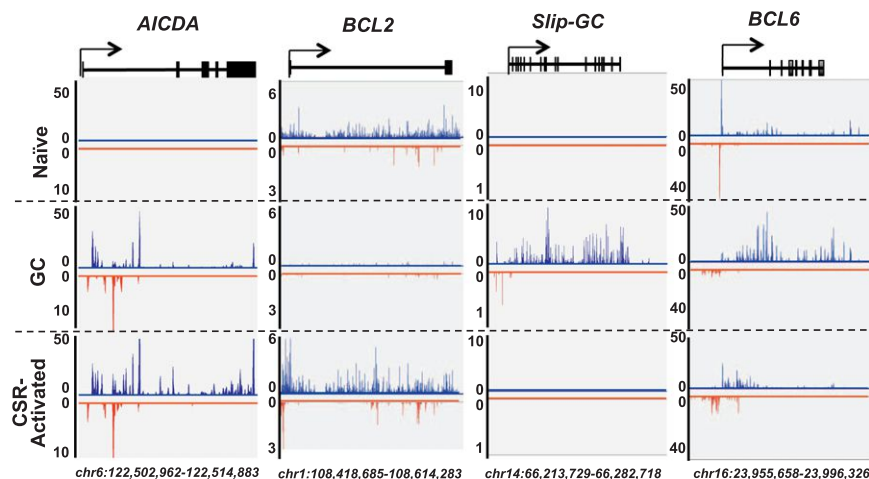


Figure 1. GRO-Seq Profiles of Naive, Germinal Center, and CSR-Activated B Cells

GRO-seq profiles of four representative genes are shown for different B cell types. The y axis indicates GRO-seq counts normalized to number of reads per million. Gene sense and antisense transcription are displayed in blue and red, respectively. Gene exons are illustrated by squares along gene bodies at the top of each panel. Arrows indicate TSSs and direction of sense transcription. Genome coordinates (mm9/NCBI37) are labeled at the bottom. All the profiles were generated from merged data of three independent experiments, which individually showed similar patterns.

See also Figure S1 and Table S1.

in naive B cells (Figure 1). In contrast, several GC B cell-specific genes, including *SLIP-GC* (Richter et al., 2009) and *Bcl6* (Basso and Dalla-Favera, 2010), had high sense transcription through their gene bodies in GC B cells, but not in naive or CSR-activated B cells (Figure 1). Finally, *Bcl2*, which is expressed in CSR-activated but not in GC B cells (Liu et al., 1991), showed corresponding sense transcription patterns (Figure 1).

While *IgH C_H* exons were appropriately transcribed in the three cell populations (Figure S1H), transcription within core S regions could not be mapped due to their abundant repetitive sequence (Pavri et al., 2010). All analyzed mice had a clonal knock-in *V_H(D)* *J_H* exon (*V_HB1-8*) (Sonoda et al., 1997), which showed active transcription at its upstream regions in all three cell types (Figure S1H). However, detailed analyses of transcription through the body of the *V_HB1-8* allele was not possible (Figure S1H); because it uses a member of the *V_HJ558* family, which contains many highly related, unexpressed upstream copies (Brodeur and Riblet, 1984).

Enhanced Identification of AID Off-Target Sites in α CD40 plus IL4-Stimulated B Cells

We developed high-throughput genome-wide translocation sequencing (HTGTS) to map, at the nucleotide level, translocation junctions between bait I-SceI nuclease generated DSBs in *c-myc* and other endogenous DSBs (Chiarle et al., 2011). Identification of DSB hotspots from a fixed chromosomal site is facilitated by ability of recurrent DSBs to dominate genome-wide translocation landscapes due to cellular heterogeneity in 3D genome organization (Zhang et al., 2012). Beyond expected Ig locus targets, our prior HTGTS studies revealed 15 non-Ig genes that are recurrent targets of AID-initiated DSBs and translocations (Chiarle et al., 2011) (Table S2). To increase the depth of HTGTS AID off-target data and allow better comparison with deeper GRO-seq transcription profiles, we further employed a modified, more sensitive HTGTS approach (Frock et al., 2015), coupled with ataxia telangiectasia mutated (ATM)-deficient CSR-activated B cells (Hu et al., 2014). This combined approach identified highly clustered AID-dependent off-target DSB sites within 36 additional genes (Figure S2A; Extended Experimental Procedures; Table S2). Overall, we now have identified 51 highly focal AID off-target

DSB/translocation sites in α CD40 plus IL4-stimulated B cells (Table S2). Nearly 90% of the new off-target set was validated in WT B cells by HTGTS and/or by an independent method (Qian et al., 2014 in this issue of *Cell*) (Extended Experimental Procedures). As previously found for our more limited set of AID off-target sites (Chiarle et al., 2011), many of our new AID off-targets occurred within genes that have divergently transcribed TSSs; but the focal sites of HTGTS junctions within them were downstream of and distinct from divergently transcribed TSSs (Chiarle et al., 2011) (Figure 2). Thus, we were compelled to search for other factors that promote such focal AID off-targeting. As we found no enrichment for known AID targeting motifs in these regions (Extended Experimental Procedures), we focused our search on potentially novel transcriptional and/or epigenetic features and, as described below, identified both.

AID Off-Targets Cluster at Sense/Antisense Transcription Sites Downstream of the TSS

With our present, substantially deeper, GRO-seq data sets, we further analyzed potential relationships between sense/antisense transcription and AID off-target sites in α CD40 plus IL4-activated B cells. Initially, we visually inspected three linked AID off-target sites, including sites in the previously characterized *IL4r* and *IL21r* genes (Chiarle et al., 2011) and a newly identified site in *Nsmce1*. In each of these linked genes, HTGTS translocation junctions were tightly clustered in a region downstream of the TSS (Figure 2A). Moreover, in each, translocation clusters fell within sites that exhibited enriched, overlapping sense and antisense transcription to which we heretofore apply the term “convergent transcription” (ConvT) (Figures 2A and 3A). We also found a robust AID off-target site within the AID gene (*Aicda*) itself (Figure 2B; Table S2). *Aicda* is associated with five enhancers that lie upstream, within, or downstream of the gene body (Kieffer-Kwon et al., 2013; Matthews et al., 2014) (Figure 2B). Four of these enhancers showed both sense and antisense transcription, likely at least in part in the context of generating eRNAs (Natoli and Andrau, 2012) (Figure 2B). Notably, the major focal cluster of AID off-target sites in and around *Aicda* fell within a ConvT region associated with enhancer 4 downstream of the TSS (Figure 2B).

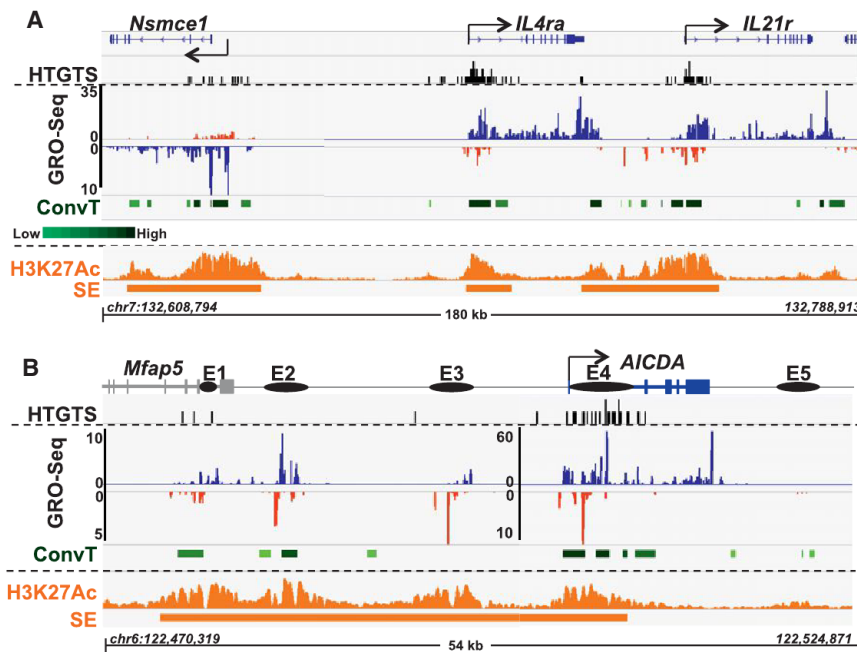


Figure 2. AID Off-Target Translocations Cluster within Regions of ConvT and SEs

(A) HTGTS, GRO-seq, ConvT, and H3K27Ac profiles in the vicinity of *Nsmce1*, *IL4ra*, and *IL21r* genes. Top: HTGTS junctions are indicated by black bars. Middle (GRO-seq): GRO-seq-determined sense and antisense transcription is displayed in blue and red, respectively. ConvT regions are shown as green bars at the bottom with the darkest shades corresponding to highest levels of ConvT as calculated by the geometric means of sense and antisense transcription reads (see Extended Experimental Procedures). A scale bar is shown below the ConvT label. Bottom (H3K27Ac and SE): the H3K27Ac ChIP-seq profile is shown in orange and identified SEs depicted below with orange bars. *Nsmce1* TSS is manually curated based on GRO-seq profile.

(B) Profile of *AICDA* gene. Known *AICDA* enhancers are represented as E1–E5 with solid circles. To represent lower level transcription of certain enhancers, a smaller scale is used for E1–E3. Genome coordinates (mm9/NCBI37) are at the bottom of each panel. Other details are the same as for (A).

See also Figure S2 and Table S2.

Genome-wide Association of ConvT and AID Off-Targets in CSR-Activated B Cells

Visual inspection of AID off-target sites in additional genes revealed similar coincidence of regions of robust sense/antisense (S/AS) ConvT downstream of the TSS with focal clusters of AID-dependent off-target translocations (see below, Figure S2), leading us to examine this potentially striking association genome-wide. While metagene profiles of GRO-seq data from α CD40 plus IL4-activated B cells confirmed expected DivT at many TSSs (Wu and Sharp, 2013), they did not reveal similarly abundant convergent transcription (Figure S1F). Thus, at least at robust levels, convergent transcription likely occurs in a much smaller fraction of genes (Figure S1F). For further analyses, we developed a computational pipeline to specifically identify S/AS ConvT regions genome-wide using deep GRO-seq data sets (Figure 3A; Extended Experimental Procedures). Strikingly, among the 51 AID off-target genes, 48 (94%) had their highly clustered AID off-target translocations within regions associated with S/AS convergent transcription (Figure 3B). We randomly sampled convergent transcription of regions, in the top three transcription-level deciles, that were similar in size to those of AID off-target regions and found a much lower association with convergent transcription than for AID off-target regions (Figure S3A). This finding shows that AID off-targets are highly enriched at ConvT sites. Finally, concurrency between S/AS convergent transcription and AID off-target translocations was much higher in α CD40 plus IL4-activated B cells (94%) than in naive (49%) or GC (63%) B cells, consistent the notion that not all AID off-targets would be shared among three cell types with overlapping, but clearly distinct, transcription profiles (Figures 3B and S3B; also see below).

To further examine the relationship between ConvT and AID targeting, we calculated the geometric mean of GRO-seq sense

and antisense transcription reads in regions of interest to quantify degree of convergent transcription (Extended Experimental Procedures) and divided the values into deciles displayed by different shades of green bars below the GRO-seq profiles (Figures 2 and S2; dark green is highest and light green lowest levels). For most AID off-targets, HTGTS junctions clustered in regions with the most abundant ConvT (Figures 2 and S2). Furthermore, ConvT associated with AID off-targets was substantially greater than that at other genomic loci (Figure S3C). In addition, within AID off-target ConvT regions, the highest density of translocations occurred at sites with the most robust ConvT (Figure 3C). We further evaluated this relationship by determining how variations in sequencing depth influenced identification of ConvT. Even with our current very deep sequencing depth (>306 million mappable reads), we did not reach saturation of the total length of ConvT regions (Figure S3D), consistent with (at least low-level) pervasive transcription of the genome (Jacquier, 2009). In contrast, we reached saturation of the concurrency of AID off-targets with ConvT regions at \sim 40% of our current GRO-seq depth (120 million mappable reads; Figure S3D), confirming that most AID off-target DSB/translocation regions detectable by HTGTS in α CD40 plus IL4-stimulated B cells are associated with relatively strong convergent transcription (Figures 3C and S3D).

Convergent Transcription at AID Off-Targets Arises from Intragenic SEs

ConvT of overlapping genes was first described in bacteriophage lambda (Ward and Murray, 1979) and has been associated with transcriptional gene silencing (Gullerova and Proudfoot, 2012) and RNA Pol II collision (Hobson et al., 2012). Considering that intragenic antisense transcription associated with AID-off target sequences may arise from enhancer elements, we explored whether intragenic SEs were enriched for

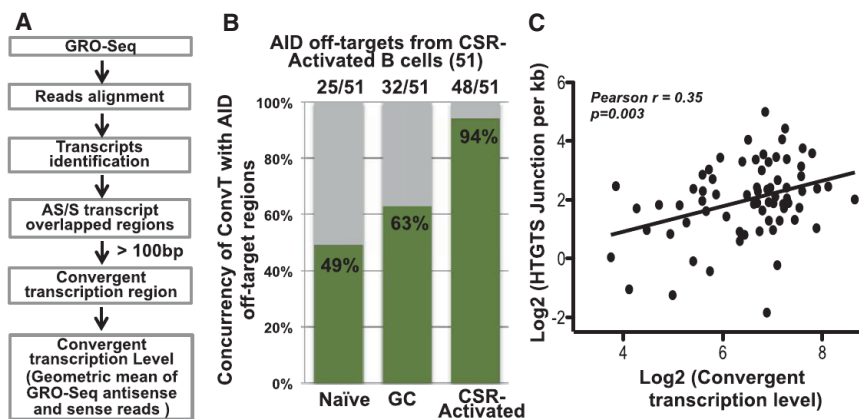


Figure 3. AID Off-Targets Correlate with ConvT in CSR-Activated B Cells

(A) Pipeline for identification of ConvT regions. Raw GRO-seq reads were aligned to the genome and transcripts were identified de novo. A “ConvT” region was defined as sense and antisense transcription overlaps that were longer than 100 bp. See Extended Experimental Procedures for details. (B) The percentage of the 51 AID off-target regions identified in CSR-activated B cells that were associated with ConvT regions in the three listed cell populations is indicated by the green bars. (C) Numbers of translocation junctions per kb (y axis) plotted against ConvT levels (x axis) of all individual AID off-target regions except *Pvt1* (see Extended Experimental Procedures). Pearson's correlation coefficient and two-tailed p value are indicated. See also Figure S3.

AID off-targets compared to typical enhancers. Enhancer regions were identified by triplicate chromatin immunoprecipitation sequencing (ChIP-seq) using an antibody to the active enhancer histone mark H3K27Ac in chromatin purified from α CD40 plus IL4-stimulated B cells (Figure S4A). SEs were called for regions of asymmetric, high enrichment for H3K27Ac, as previously described (Whyte et al., 2013). We found the *Aicda* locus to be largely encompassed within a SE in CSR-activated B cells with robust H3K27Ac signals over E1, E2, E3, and E4 (Figure 2B), the active enhancers in CSR-activated B cells (Kieffer-Kwon et al., 2013; Matthews et al., 2014). Notably, E4 also corresponds in position to a cluster of HTGTS junctions and robust ConvT (Figure 2B). Likewise, the *Nsmce1*, *IL4ra*, *Il21r*, and many other AID off-target genes were each associated with SEs and again the peak of HTGTS junctions and regions of robust ConvT occurred within regions of robust H3K27Ac SE signals (Figures 2A and S2).

We performed an unbiased association analysis between the 51 AID off-targets identified by HTGTS and the non-Ig 448 SEs that we identified in α CD40 plus IL4-activated B cells. These studies revealed that 50 of the 51 AID off-target genes in these cells are associated with SEs and that the discrete translocation clusters were within SEs (Figure 4A). Notably, the single AID off-target region not within a SE (under the current cutoff for SE identification; Extended Experimental Procedures) was in a typical enhancer (Table S2). In addition, 47 (92%) of the AID off-target translocation clusters were within regions of SEs that overlap with annotated gene bodies (Figure 4A). The other three HTGTS off-target translocation clusters occurred within transcribed regions of SEs that have not yet been assigned to a target gene (Table S3). As a comparison, random samplings of transcribed genomic regions corresponding in size to those of AID off-targets yielded at most three (6%) that overlapped with SEs. Independent analysis of the relationship between HTGTS hotspots and H3K27Ac ChIP-seq using an orthogonal computational method identified 41 AID off-targets within SE domains (Figures S4C and S4D), including additional off-targets that correlated with robust ConvT (Figure S2; Table S2; Extended Experimental Procedures). Finally, within a given AID off-target region, translocation junction frequency highly correlated with H3K27Ac abun-

dance (Figure S4B). In this regard, SEs associated with AID off-target sequences were more enriched for H3K27Ac, compared to other SEs (Figure 4B). Thus, the relative activity of SEs, estimated by regional histone acetylation, correlates with the frequency of AID off-targets within them.

The majority (30 of 51) of the AID off-target genes had a SE that overlapped with the region just downstream of the TSS that was enriched in AID off-targets, as represented by the *CD83* gene (Figures 4C and S2). In addition, a number (12 of 51) of the AID targets were relatively small genes, such as *Pim1*, that were located within large SEs and, correspondingly, off-target translocations tended to span the gene body (Figures 4D and S2). Several AID off-target genes (3 of 51) were large genes, such as *Pvt1*, the well-known translocation target downstream of *c-myc*, in which translocations clustered within SEs that occurred inside the gene body (Figure 4E). Finally, the remainder (6 of 51) fell into a heterogeneous set in which AID off-target translocations clustered into convergently transcribed SE domains that, for various reasons were not yet assignable to a specific gene (e.g., *Gpr183*; Figure S2C).

Intragenic SEs with Robust ConvT Represent the Most Common AID Off-Targets

Nearly all AID off-target clusters identified by HTGTS in α CD40 plus IL4-activated B cells are associated with SEs; yet, only a subset of SEs are AID off-targets. Motivated by the putative contribution of S/AS eRNA transcription to translocation frequency, we compared regions of AID off-target genes where SEs overlap with the gene body (intragenic SEs) to regions where SEs lie outside the gene body (intergenic SEs) and to regions of gene bodies that do not overlap with SEs (nonoverlapping gene region), for translocation density (translocations per 1 kb; Figure 5A) and for ConvT levels (geometric means; Figure 5A). We observed that translocation junction density and ConvT levels in AID off-target regions are highly enriched among intragenic SEs compared to both intergenic SEs and nonoverlapping gene regions (Figure 5A; upper). Despite this enrichment, only ~10% of all intragenic SEs in the CSR-activated B cells are AID off-targets (Figure 4A; Table S3) and other SE-gene overlap regions exist that are not enriched in AID off-target activity

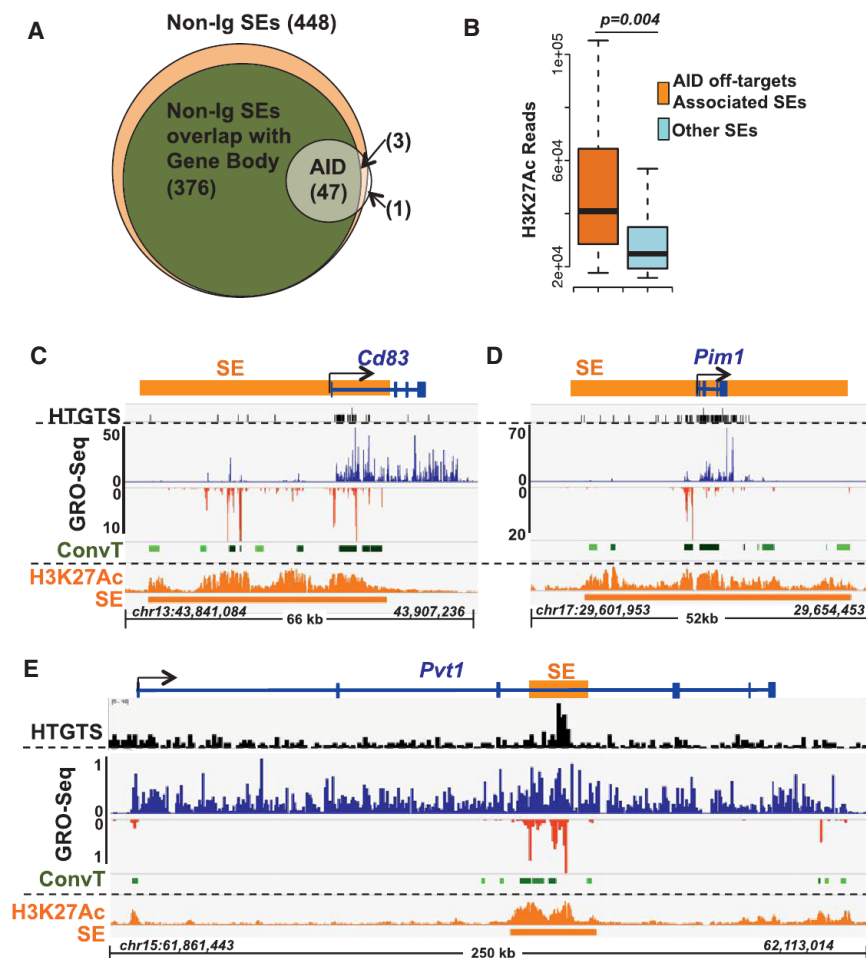


Figure 4. AID Off-Target ConvT Arising from Intragenic SEs

(A) Venn diagram showing the number of AID off-target regions that overlapped with total non-Ig SEs (448) and with non-Ig SEs overlapping with Gene Bodies (376).

(B) H3K27Ac signals of AID off-target-associated SEs (orange) and the other SEs (cyan) are plotted. AID off-target-associated SEs had a stronger H3K27Ac signal (Mann-Whitney U test, p value = 0.004).

For C, D, and E, representative AID off-targets are shown based on the SE location indicated in the diagram at the top of each panel.

(C) Many AID targets locate downstream of TSSs where SEs and genes overlap. *CD83* is shown as an example.

(D) For some relatively small genes located within a larger SE, nearly the whole gene body is an AID off-target, as shown for *Pim1*.

(E) SEs inside of very long genes, like *Pvt1* also provide focal AID off-targets. HTGTS, GRO-seq, and H3K27Ac/SE data is illustrated for each panel as described in Figure 2A. The relatively high HTGTS background in *Pvt1* results from long resections downstream of the HTGTS bait DSB in *c-myc* (Chiarle et al., 2011).

See also Figure S4 and Table S3.

AID Off-Targets in GC B Cells Associate with Convergent Transcription

Prior studies of a selected set of AID off-targets divided them into three groups in GC B cells based on mutation frequency in Ung/Msh2 double-deficient B cells versus AID-deficient B cells, including 15 group A genes that had high levels of mutation,

(Figure 5B; upper). Comparison of ConvT levels in each of the three regions outlined above (Figures 5A and 5B, lower panels) revealed that intragenic SEs featuring high levels of ConvT were more frequently AID off-target regions than intragenic SEs lacking high-level S/AS transcription (Figures 5A and 5B, lower panels).

Finally, to further address why some SEs are AID targets and others are not, we grouped all intragenic SEs into deciles based on low to high convergent transcription (Figure 5C). We then calculated the percentage of the combined 228 unique AID off-targets revealed by HTGTS (this study) and by an independent RPA-ChIP study (Qian et al., 2014) in CSR-activated B cells in each decile. Strikingly, 60% of all SEs within the top two deciles (highest convergent transcription) were sites of clustered AID off-target DSBs and/or translocations. Comparative analysis of SEs in these top two deciles that were AID off-targets versus those that were not did not reveal any obvious sequence differences (e.g., GC content or WRCH and AGCT motifs density). However, ConvT regions associated with SEs in the top two deciles that were AID off-targets were significantly longer than those that were not (Figure 5D). These studies provide strong evidence that ConvT from intergenic SEs generates a major class of focal AID off-target regions.

21 group B genes that had substantially lower levels, and 47 group C genes that were infrequently mutated (Liu et al., 2008). Our GRO-seq analyses of GC B cells revealed that ~70% of the highly mutated group A gene off-target regions, including *Pim1*, *Ebf1*, *CD83*, and *Ocab*, overlapped with ConvT regions (Figures 6A and 6C) that were well above simulated background levels expected for the most highly transcribed genes (Figure S5A). In contrast, regions reported to have low level mutation frequency (groups B and C genes) showed low correlations with convergent transcription (33% and 32%, respectively; Figure 6A) that were not above simulated background concurrency (Figure S5A). Finally, of the five group A genes that did not associate directly with convergent transcription, SHMs in four occurred quite proximal to ConvT regions (Figure S5C). We identified SEs in GC B cells via H3K27Ac ChIP-seq analyses. We found that some SEs were shared between GC and CSR-activated B cells, while many others were found only in one or the other cell type (Figure S5B; Table S3), consistent with the overlapping but distinct GRO-seq profiles of these two cell types (Figure S3B; Table S1). Of the highly mutated group A gene regions, nearly half were associated with SEs (Figure 6B) and all were associated with H3K27Ac peaks (Figures 6C and S5C). For group B and C gene regions, concurrencies with SE were

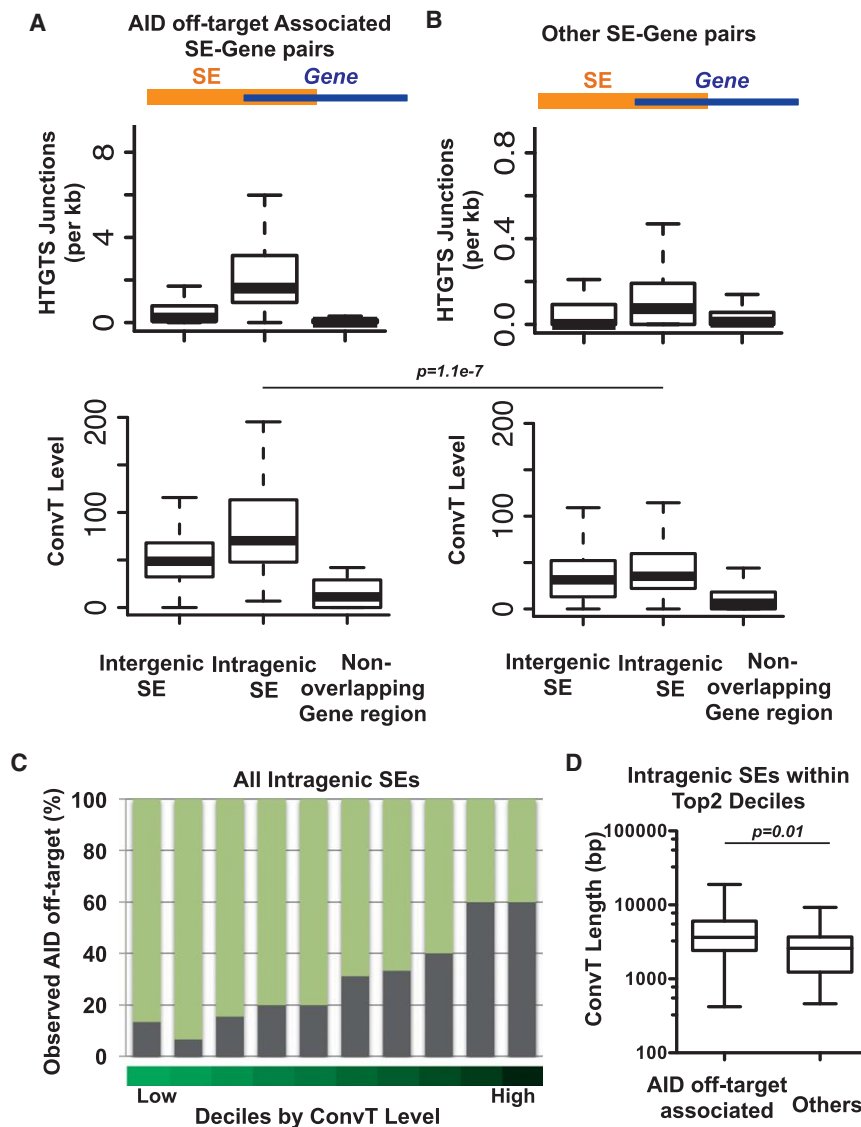


Figure 5. Convergetly Transcribed Intragenic SEs Are Preferred AID Off-Targets

(A) Upper and lower: each SE associated with an AID off-target region and its overlapping gene body were divided into intergenic SEs, intragenic SEs, and nonoverlapping gene regions as described in the text and outlined at the top of the panels. For all AID off-targets, the number of translocation junctions per kb in each of the three regions (upper panel) and convergent transcription levels of each region (lower panel) are plotted.

(B) Upper and lower: each SE that was not associated with an AID off-target region and its overlapping gene body were divided into regions as described for (A) and translocation junction numbers per kb (upper panel) and convergent transcription levels (lower panel) plotted for each region. A Mann-Whitney U test was performed to compare two classifications of SEs for convergent transcription ratios within each of the three regions; the only significant difference found was that the AID-off-target intragenic SEs has a significantly higher convergent transcription ratio than non-AID off-target intragenic SEs (p value = 1.1×10^{-7}).

(C) All intragenic SEs were grouped into deciles based on the ConvT levels. The fraction of AID off-targets in each decile is indicated by gray bar.

(D) Intragenic SEs in the top two deciles are divided into those associated with AID off-targets (60%) and those that are not (40%). Length of ConvT regions was plotted and found to be significantly longer in the AID off-target-associated intragenic SEs (Mann-Whitney U test, p value = 0.01).

20% and 2%, respectively. Thus, under physiological conditions in the GC, AID often targets convergetly transcribed intragenic SEs or, occasionally, typical enhancers.

Convergetly Transcribed Intragenic SEs Target AID in Non-B Lymphoid and Human Cells

Ectopic manipulation of endogenous SEs and ConvT regions to assess affects on AID targeting would be problematic because these regions are the actual AID targets. As an alternative approach, we performed GRO-seq on mouse embryonic fibroblasts (MEFs) in which ectopic AID expression revealed a set of 29 AID off-target sequences, most of which were novel (Qian et al., 2014) (Table S4). Remarkably, we found that the great majority of these clustered MEF translocations occurred in ConvT regions (Figures 7A and S6A) that also were mostly also associated with SEs (Qian et al., 2014) (Table S4). We also tested the generality of our ConvT findings with respect to AID

below, we have also extended our findings to human B cell lymphoma translocations.

DISCUSSION

Off-Target AID Activity in Convergetly Transcribed Intragenic SEs

We report that most AID off-target DSBs and translocations in CSR-activated B cells occur in and around ConvT regions within genes (Figure 3). Furthermore, most of these AID off-target sites in CSR-activated B cells occurred within portions of genes that overlapped with enhancers, the vast majority of which were SEs (Figure 4). Together, these findings implicate a role for SEs within genes in generating robust ConvT and, thereby, in creating susceptibility to AID off-target activity. Notably, we also found that the majority of the regions with highest levels of off-target AID activity in GC B cells or in human Ramos cells

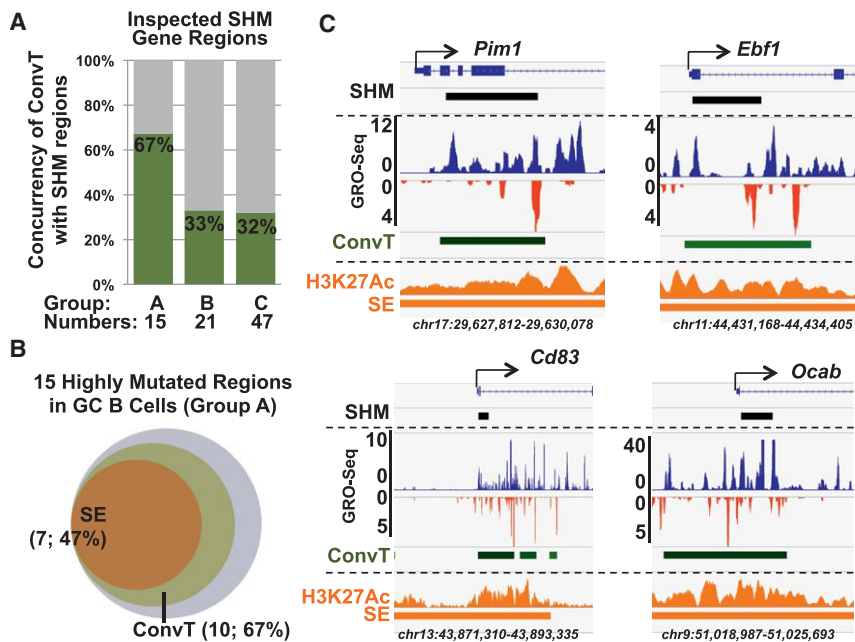


Figure 6. ConvT and SEs Correlate with AID Off-Targets in GC B Cells

Regions of genes containing SHMs in Ung/Msh2 double deficient GC B cells were analyzed for convergent transcription as determined by GRO-seq and outlined in Figure 3. GC AID off-target group A, B, and C genes include gene regions with high, intermediate, and low frequencies of AID-dependent mutations, respectively (Liu et al., 2008).

(A) Concurrency of group A, B, and C gene ConvT regions in GC B cells.

(B) Venn diagram showing the number of group A gene regions that overlapped with SEs and ConvT. (C) Examples of group A gene regions are shown. Approximately 2–3 kb regions around the TSSs of the indicated genes are shown. The “SHM” diagram at the top of each subpanel indicates regions of these genes included in the prior SHM analyses (Liu et al., 2008) with a black bar. GRO-seq profile, ConvT, H3K27Ac ChIP-seq profile, and SEs are shown as in Figure 2A.

See also Figure S5 and Table S3.

undergoing SHM are in focal areas of target genes that contain SEs and undergo robust ConvT (Figures 6 and 7). Even in non-lymphoid cells (MEFs) in which AID was ectopically expressed, we found that the great majority of 29 AID-dependent translocation clusters occurred in regions that underwent robust ConvT (Figure 7), confirming our findings for a totally different set of genes in a different cell type. Together, these findings strongly support a mechanistic link between AID off-target sequences and S/AS convergent transcription. A role for SEs in AID off-targeting also has been revealed by a separate study (Qian et al., 2014).

Potential Mechanisms by which SEs and ConvT Contribute to AID Off-Target Activity

RNA polymerase II (Pol II) transcriptional pausing or stalling contributes to directing AID to Ig gene SHM and CSR targets via a process thought to involve AID association with the Spt5 transcription cofactor (Pavri et al., 2010; Storb, 2014). Ig gene V(D)J exons and S regions likely evolved specific features to promote AID targeting (Alt et al., 2013). As AID off-target genes lack consistent sequence features of Ig gene AID targets (Duke et al., 2013), the question of how they attract AID has been long-standing. Our current findings implicate a mechanism that answers this question for the majority of AID off-targets (Figure 7C). Thus, most robust AID off-target DSBs, SHMs, and translocations occur within intragenic SEs, where we find ConvT that includes sense gene transcription and antisense transcription emanating from the SEs. In such AID off-target regions, antisense eRNA transcription generally occurs at lower levels than sense transcription (Figures 2 and 4). Thus, most genic sense transcription likely proceeds unimpeded to generate full length mRNAs with only a small fraction encountering antisense transcription, consistent with ability of cells to generate products of these genes (Storb, 2014). Prior yeast studies showed that,

within convergently transcribed sequences, Pol II elongation complexes proceeding in opposite directions cannot bypass each other, and consequential Pol II collisions lead to stalling or stopping (Hobson et al., 2012). We propose that such Pol II stalling due to convergent transcription leads to AID recruitment and further downstream events similar to those implicated in specialized Ig gene targets (Figure 7C) (Pavri et al., 2010; Basu et al., 2011). Beyond AID recruitment, convergent transcription could also generate ssDNA substrates for AID. Thus, following Pol II collisions, RNA exosome or other RNase activities could remove nascent transcripts (Basu et al., 2011; Pefanis et al., 2014; Andersson et al., 2014) to provide local ssDNA targets (Figure 7C).

Implications of AID Off-Target Activity for AID On-Target Ig Gene Activity

AID activity generally occurs at much higher levels on specialized Ig gene targets than on off-targets (Liu and Schatz, 2009; Yamane et al., 2011; Chiarle et al., 2011; Klein et al., 2011). Whether or not the ConvT mechanism we propose for off-targets can be applied to on-targets remains to be determined. In CSR-activated B cells, we observed ConvT within the very 5' S_μ region (Figure S1H). However, the transcription profile of core S regions cannot be obtained due to poor mappability of repetitive S regions (Pavri et al., 2010). Clearly, S regions evolved specialized structural features that facilitate AID recruitment and access to the ssDNA substrates (Alt et al., 2013). However, mechanisms by which AID specifically targets Ig variable region exons for SHM in GCs may be more relevant. In this regard, a long-standing paradox involves that fact that SHM of variable region exons occurs only in GC B cells and not in CSR-activated B cells, even though the variable region exons are transcribed in both (Liu and Schatz, 2009). Our preliminary analyses reveal potentially higher relative levels of

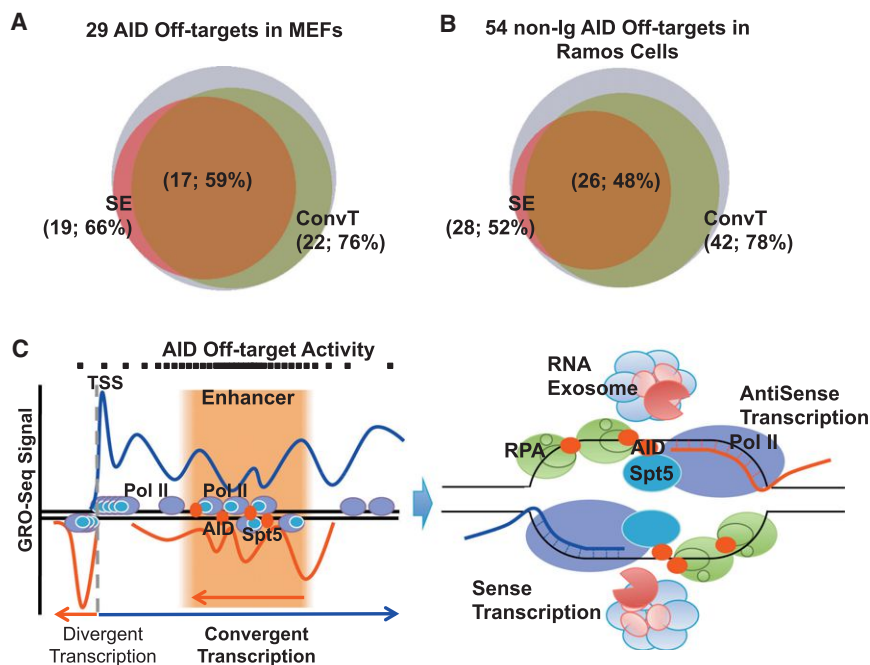


Figure 7. Model of AID Targeting at Off-Targets

(A) Venn diagram showing the number of AID off-target regions that overlapped with SEs and ConvT in MEFs with ectopic AID overexpression.

(B) Venn diagram showing number of AID off-target regions that overlapped with SEs and ConvT in Ramos Human Burkitt's lymphoma cell line.

(C) Model of AID "off-targeting." Left: at AID off-targets, SEs overlap with gene bodies and this combination generates regions of sense/antisense convergent transcription due to sense gene transcription encountering the enhancer antisense transcription. Right: stalled RNA polymerase with the help of Spt5 recruits AID and generates regions of ssDNA. RNA Exosome or other RNases degrade the aborted sense and antisense transcripts and works together with RPA to help AID access to the ssDNA substrates. Some aspects adapted from Basu et al. (2011). See Discussion for other details. See also Figure S6 and Table S4.

antisense to sense transcription on the downstream edge of the KI V(D)J (VB1-8) exon in GC versus naive or CSR-activated B cells (Figure S1H). However, as we cannot map transcription within the main body of the KI VB1-8 due to many highly related unexpressed, upstream V_HJ558 sequences, final testing of this potential mechanism for specific AID targeting of V(D)J exons will require additional mouse models that eliminate sequence redundancies.

Role of SE Transcription in Genome Instability and Cancer

SEs are important for establishment of cell lineage and expression of cell lineage-specific genes (Whyte et al., 2013; Hnisz et al., 2013). Correspondingly, SEs are associated frequently with genes highly expressed in activated B cells (Table S3). Many of the 51 genes that we have shown to have SEs that are AID off-targets are B cell-specific genes and a notably high proportion (25%) are known oncogenes (Figure S2B). In this regard, many human B cell lymphomas contain translocations or mutations of oncogenes that are initiated by off-target AID activity (Alt et al., 2013; Küppers and Dalla Favera, 2001). Reminiscent of the AID off-targeting pattern in mouse CSR-activated and GC B cells, human B cell oncogene translocation sites often occur downstream of TSSs (Migliazza et al., 1995; Pasqualucci et al., 2001; Shen et al., 1998). Indeed, we have analyzed SEs in human tonsil B cells (enriched in GC B cells) and found that many oncogene translocations in human B cell lymphoma, including those in *c-myc*, *Pax5*, *Bcl6*, *Bcl2*, *Pim1*, *Ocab*, *Lcp*, and *Bcl7a*, occur in regions downstream of TSSs where SEs overlap with gene bodies (Figure S6C). Thus, beyond contributing to deregulated oncogene expression (Chapuy et al., 2013), our findings suggest that SEs may target oncogenes for translocations in B cell lymphoma. Finally, AID also

has been implicated in genomic instability and translocations in cells beyond those of the immune system (Lin et al., 2009; Marusawa et al., 2011). Our MEF studies suggest ConvT from SEs could play a role in such settings.

EXPERIMENTAL PROCEDURES

B Cell Purification

Splenic naive B cells were purified from V_HB1-8 heavy chain knock-in mice as described (Cato et al., 2011). Naive B cells were activated with α CD40 plus IL4 for 60 hr to generate CSR-activated B cells. V_HB1-8 knock-in mice were immunized with 5×10^8 sheep red blood cells (SRBCs) for 9 days. Splenic GC B cells were purified as described (Cato et al., 2011) (see Extended Experimental Procedures for details.) All animal experiments were performed under protocols approved by the Institutional Animal Care and Use Committee of Boston Children's Hospital.

GRO-Seq and ChIP-Seq

GRO-seq (Core et al., 2008) and H3K27Ac ChIP-seq (Chapuy et al., 2013) were performed as described. Three biological replicates of each mouse B cell type were performed. Two biological replicates of mouse MEF experiments and one biological replicate of Ramos experiments were performed.

AID Off-Targets

HTGTS was performed with α CD40 plus IL4 or RP105-activated ATM-deficient CSR-activated B cells as described (Hu et al., 2014) and also with a new HTGTS method (Frock et al., 2015). AID off-target coordinates were retrieved via a new HTGTS pipeline (Frock et al., 2015) (see Extended Experimental Procedures for details).

Data Analysis

GRO-seq and ChIP-seq data sets were aligned using Bowtie software (Langmead and Salzberg, 2012) to mouse genome build mm9/NCBI37 or human genome build hg19/NCBI37. Uniquely mapped, nonredundant sequence reads were retained. We used Homer software (Heinz et al., 2010) to de novo identify transcripts from both strands of the genome in the context of the GRO-seq data and considered broad sense/antisense overlap regions (>100 bp) as ConvT regions. We used the MACS1.4 software (Zhang et al., 2008) to identify regions of ChIP-seq enrichment over background with a p value threshold of 10^{-5} . We used ROSE software to

identify SEs (Whyte et al., 2013) (see Extended Experimental Procedures for details).

ACCESSION NUMBERS

The Gene Expression Omnibus databank accession number for all deep sequencing data reported in this paper is GSE62296.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.014>.

AUTHOR CONTRIBUTIONS

F.L.M., Z.D., A.F., J.E.B., X.S.L., and F.W.A. designed the study. F.L.M., C.A., and C.R.W. purified B cells and performed GRO-seq. Z.D. analyzed GRO-seq data. F.L.M., A.F., and C.A. performed H3K27Ac ChIP-seq. Z.D. and A.F. analyzed ChIP-seq data. J.H. performed HTGTS. J.H. and R.M. analyzed HTGTS data. Q.W. and M.C.N. prepared MEFs and supplied the AID off-target and SE list for MEFs. K.K. and R.C. supplied the AID off-target and SE list for Ramos cells and shared their AID off-target list for CSR-activated B cells. D.N. advised on statistical analysis and various aspects of data analysis. F.L.M., Z.D., A.F., and F.W.A. designed figures. F.W.A. and F.L.M. drafted the manuscript. F.L.M., Z.D., A.F., J.E.B., X.S.L., and F.W.A. polished the manuscript.

ACKNOWLEDGMENTS

F.W.A. was supported by NIH grants R01AI077595 and P01CA109901, and a Leukemia and Lymphoma Society (LLS) SCOR grant, and he is an investigator of the Howard Hughes Medical Institute. X.S.L. was supported by NIH grant 1R01GM099409. A.F. and J.E.B. were supported by a Leukemia and Lymphoma Society SCOR, the National Science Foundation, and NIH grants 1R01 CA176745-01 and P01 CA109901. M.C.N. was supported by NIH grants AI072529 and AI037526 and is an investigator of the Howard Hughes Medical Institute. K.K. and R.C. were supported by the intramural research program of NIAMS, NIH. F.L.M. is a Lymphoma Research Foundation postdoctoral fellow and was a Cancer Research Institute postdoctoral fellow. J.Z. is supported by a Robertson Foundation/Cancer Research Institute Irvington Fellowship. Z.D. was supported by the National Science Foundation of China grant NSFC 31329003. The authors thank Drs. Yi Zhang and Li Shen for assistance with DNA sequencing and Dr. David Schatz (Yale University) for providing primers for GC SHM targets. Dr. Bradner is a scientific founder of Syros Pharmaceuticals, which is developing drugs targeting super-enhancers.

Received: August 1, 2014

Revised: October 1, 2014

Accepted: October 27, 2014

Published: December 4, 2014

REFERENCES

- Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–731.
- Alt, F.W., Zhang, Y., Meng, F.L., Guo, C., and Schwer, B. (2013). Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell* **152**, 417–429.
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461.
- Basso, K., and Dalla-Favera, R. (2010). BCL6: master regulator of the germinal center reaction and key oncogene in B cell lymphomagenesis. *Adv. Immunol.* **105**, 193–210.
- Basu, U., Meng, F.L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R., et al. (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* **144**, 353–363.
- Brodeur, P.H., and Riblet, R. (1984). The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur. J. Immunol.* **14**, 922–930.
- Cato, M.H., Yau, I.W., and Rickert, R.C. (2011). Magnetic-based purification of untouched mouse germinal center B cells for ex vivo manipulation and biochemical analysis. *Nat. Protoc.* **6**, 953–960.
- Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G., et al. (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790.
- Chiarle, R., Zhang, Y., Frock, R.L., Lewis, S.M., Molin, B., Ho, Y.J., Myers, D.R., Choi, V.W., Compagno, M., Malkin, D.J., et al. (2011). Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107–119.
- Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936.
- Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastritis, E., Gilpatrick, T., Paranal, R.M., Qi, J., et al. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904–917.
- Di Noia, J.M., and Neuberger, M.S. (2007). Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22.
- Duke, J.L., Liu, M., Yaari, G., Khalil, A.M., Tomayko, M.M., Shlomchik, M.J., Schatz, D.G., and Kleinstein, S.H. (2013). Multiple transcription factor binding sites predict AID targeting in non-Ig genes. *J. Immunol.* **190**, 3878–3888.
- Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.-J., Kii, E., and Alt, F.W. (2014). Genome-wide detection of specific and non-specific DNA DSB activities of custom nucleases in human cells. *Nat. Biotechnol.* Published online January 4, 2015. <http://dx.doi.org/10.1038/nbt.3101>.
- Gostissa, M., Yan, C.T., Bianco, J.M., Cogné, M., Pinaud, E., and Alt, F.W. (2009). Long-range oncogenic activation of Igh-c-myc translocations by the Igh 3' regulatory region. *Nature* **462**, 803–807.
- Gullerova, M., and Proudfoot, N.J. (2012). Convergent transcription induces transcriptional gene silencing in fission yeast and mammalian cells. *Nat. Struct. Mol. Biol.* **19**, 1193–1201.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947.
- Hobson, D.J., Wei, W., Steinmetz, L.M., and Svejstrup, J.Q. (2012). RNA polymerase II collision interrupts convergent transcription. *Mol. Cell* **48**, 365–374.
- Hu, J., Tepsuporn, S., Meyers, R.M., Gostissa, M., and Alt, F.W. (2014). Developmental propagation of V(D)J recombination-associated DNA breaks and translocations in mature B cells via dicentric chromosomes. *Proc. Natl. Acad. Sci. USA* **111**, 10269–10274.
- Jacquier, A. (2009). The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.* **10**, 833–844.
- Kieffer-Kwon, K.R., Tang, Z., Mathe, E., Qian, J., Sung, M.H., Li, G., Resch, W., Baek, S., Pruet, N., Grøntved, L., et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* **155**, 1507–1520.

- Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* **147**, 95–106.
- Küppers, R., and Dalla-Favera, R. (2001). Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* **20**, 5580–5594.
- Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. *Trends Biochem. Sci.* **39**, 170–182.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.
- Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13–25.
- Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B.G., Ohgi, K., Zhang, J., Rose, D.W., Fu, X.D., Glass, C.K., and Rosenfeld, M.G. (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* **139**, 1069–1083.
- Liu, M., and Schatz, D.G. (2009). Balancing AID and DNA repair during somatic hypermutation. *Trends Immunol.* **30**, 173–181.
- Liu, Y.J., Mason, D.Y., Johnson, G.D., Abbot, S., Gregory, C.D., Hardie, D.L., Gordon, J., and MacLennan, I.C. (1991). Germinal center cells express bcl-2 protein after activation by signals which prevent their entry into apoptosis. *Eur. J. Immunol.* **21**, 1905–1910.
- Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. *Nature* **451**, 841–845.
- Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334.
- Marusawa, H., Takai, A., and Chiba, T. (2011). Role of activation-induced cytidine deaminase in inflammation-associated cancer development. *Adv. Immunol.* **111**, 109–141.
- Matthews, A.J., Zheng, S., DiMenna, L.J., and Chaudhuri, J. (2014). Regulation of immunoglobulin class-switch recombination: choreography of noncoding transcription, targeted DNA deamination, and long-range DNA repair. *Adv. Immunol.* **122**, 1–57.
- Migliazza, A., Martinotti, S., Chen, W., Fusco, C., Ye, B.H., Knowles, D.M., Offit, K., Chaganti, R.S., and Dalla-Favera, R. (1995). Frequent somatic hypermutation of the 5' noncoding region of the BCL6 gene in B-cell lymphoma. *Proc. Natl. Acad. Sci. USA* **92**, 12520–12524.
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563.
- Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* **46**, 1–19.
- Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. USA* **110**, 17921–17926.
- Pasqualucci, L., Neumeister, P., Goossens, T., Nanjangud, G., Chaganti, R.S., Küppers, R., and Dalla-Favera, R. (2001). Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* **412**, 341–346.
- Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., et al. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* **143**, 122–133.
- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* **514**, 389–393.
- Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K., Resch, W., Liang, G., Tang, Z., Mathé, E., Benner, C., et al. (2014). B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* **159**. Published online December 4, 2014. <http://dx.doi.org/10.1016/j.cell.2014.11.013>.
- Rajagopal, D., Maul, R.W., Ghosh, A., Chakraborty, T., Khamlichi, A.A., Sen, R., and Gearhart, P.J. (2009). Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. *J. Exp. Med.* **206**, 1237–1244.
- Richter, K., Brar, S., Ray, M., Pisitkun, P., Bolland, S., Verkoczy, L., and Diaz, M. (2009). Speckled-like pattern in the germinal center (SLIP-GC), a nuclear GTPase expressed in activation-induced deaminase-expressing lymphomas and germinal center B cells. *J. Biol. Chem.* **284**, 30652–30661.
- Shen, H.M., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science* **280**, 1750–1752.
- Sonoda, E., Pewzner-Jung, Y., Schwes, S., Taki, S., Jung, S., Eilat, D., and Rajewsky, K. (1997). B cell development under the condition of allelic inclusion. *Immunity* **6**, 225–233.
- Storb, U. (2014). Why does somatic hypermutation by AID require transcription of its target genes? *Adv. Immunol.* **122**, 253–277.
- Victoria, G.D., and Nussenzweig, M.C. (2012). Germinal centers. *Annu. Rev. Immunol.* **30**, 429–457.
- Wang, L., Wuerffel, R., Feldman, S., Khamlichi, A.A., and Kenter, A.L. (2009). S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *J. Exp. Med.* **206**, 1817–1830.
- Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390–394.
- Ward, D.F., and Murray, N.E. (1979). Convergent transcription in bacteriophage lambda: interference with gene expression. *J. Mol. Biol.* **133**, 249–266.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319.
- Wu, X., and Sharp, P.A. (2013). Divergent transcription: a driving force for new gene origination? *Cell* **155**, 990–996.
- Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat. Immunol.* **12**, 62–69.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.
- Zhang, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908–921.

Apoptotic Caspases Suppress mtDNA-Induced STING-Mediated Type I IFN Production

Michael J. White,^{1,5,*} Kate McArthur,^{1,5} Donald Metcalf,^{2,5} Rachael M. Lane,¹ John C. Cambier,⁹ Marco J. Herold,^{4,5} Mark F. van Delft,^{2,5} Sammy Bedoui,⁶ Guillaume Lessene,^{1,5,7} Matthew E. Ritchie,^{3,5,8} David C.S. Huang,^{2,5} and Benjamin T. Kile^{1,5,*}

¹ACRF Chemical Biology Division

²Cancer and Haematology Division

³Molecular Medicine Division

⁴Molecular Genetics of Cancer Division

The Walter and Eliza Hall Institute of Medical Research, Parkville 3052, Australia

⁵Department of Medical Biology

⁶Department of Microbiology and Immunology

⁷Department of Pharmacology and Therapeutics

⁸Department of Mathematics and Statistics

The University of Melbourne, Parkville 3010, Australia

⁹Integrated Department of Immunology, University of Colorado Denver School of Medicine and National Jewish Health, Denver, CO 80206, USA

*Correspondence: mwhite@wehi.edu.au (M.J.W.), kile@wehi.edu.au (B.T.K.)

<http://dx.doi.org/10.1016/j.cell.2014.11.036>

SUMMARY

Activated caspases are a hallmark of apoptosis induced by the intrinsic pathway, but they are dispensable for cell death and the apoptotic clearance of cells in vivo. This has led to the suggestion that caspases are activated not just to kill but to prevent dying cells from triggering a host immune response. Here, we show that the caspase cascade suppresses type I interferon (IFN) production by cells undergoing Bak/Bax-mediated apoptosis. Bak and Bax trigger the release of mitochondrial DNA. This is recognized by the cGAS/STING-dependent DNA sensing pathway, which initiates IFN production. Activated caspases attenuate this response. Pharmacological caspase inhibition or genetic deletion of caspase-9, Apaf-1, or caspase-3/7 causes dying cells to secrete IFN- β . In vivo, this precipitates an elevation in IFN- β levels and consequent hematopoietic stem cell dysfunction, which is corrected by loss of Bak and Bax. Thus, the apoptotic caspase cascade functions to render mitochondrial apoptosis immunologically silent.

INTRODUCTION

Caspases are a family of 12 cysteinyl aspartate-specific proteases traditionally classified as inflammatory or apoptotic (McIlwain et al., 2013). Inflammatory caspases (caspase-1, -4, -5, and -12 in humans) mediate innate immune responses by cleaving precursors of proinflammatory cytokines such as IL-

1 β and IL-18, thereby facilitating their secretion. The apoptotic caspases (caspase-3, -6, -7, -8, and -9) play a role in the regulation of programmed cell death.

Apoptosis comprises two convergent pathways: the intrinsic and extrinsic (Youle and Strasser, 2008). The intrinsic pathway is controlled by the BCL-2 family of proteins, which is divided into three groups. The first contains prodeath BAK and BAX, the essential effectors of the pathway. Second are the pro-survival proteins (BCL-2, BCL-X_L, BCL-W, MCL-1, and A1), whose function is to prevent activation of BAK and BAX by physically restraining them and by sequestering a third group of BCL-2 family members, the prodeath “BH3-only” proteins (e.g., BIM and BID). In a healthy cell, prosurvival proteins keep BAK and BAX in check. Apoptotic signals trigger the BH3-only proteins to activate BAK/BAX. The latter induce mitochondrial outer-membrane permeabilization (MOMP), facilitating the efflux of factors, including cytochrome c, into the cytoplasm. Cytochrome c forms the apoptosome complex with APAF-1 and the inactive zymogen of the initiator caspase, caspase-9. This results in the activation of caspase-9, which then triggers the rest of the caspase cascade, culminating in activation of the effector caspases, caspase-3 and caspase-7.

The purpose of the caspase cascade remains an enigma. It mediates many of the hallmarks of apoptosis in vitro, such as DNA fragmentation and phosphatidylserine (PS) exposure, but is largely dispensable for the apoptotic death and clearance of cells in vivo. The hematopoietic system is a good example: *Bak*^{-/-} *Bax*^{-/-} mice exhibit a massive accumulation of mature blood cells, whereas mice with an *Apaf-1*^{-/-}, *Casp9*^{-/-}, or *Casp3*^{-/-} *Casp7*^{-/-} hematopoietic system show no significant perturbations in blood cell number (Lakhani et al., 2006; Lindsten et al., 2000; Marsden et al., 2002). This dichotomy can be

explained by the fact that the “point of no return” in apoptosis is BAK/BAX-mediated mitochondrial damage. Cells lacking BAK and BAX are resistant to a wide range of apoptotic stimuli; they do not exhibit cytochrome c release or caspase activation and are able to maintain clonogenicity (i.e., they can survive and generate viable progeny) (Lindsten et al., 2000; Wei et al., 2001). In contrast, Apaf-1- or caspase-deficient cells exhibit only short-term resistance to apoptotic stimuli and do not retain clonogenic potential (Ekert et al., 2004; Marsden et al., 2002; van Delft et al., 2010). Thus, although clearly capable of accelerating apoptosis, these and many other studies indicate that the apoptotic caspase cascade is not required for death to occur.

This raises important questions as to why caspase-deficient mice exhibit phenotypic abnormalities. For example, loss of Apaf-1, caspase-9, or caspase-3 results in lethality associated with large ectopic cell masses in the forebrain (Kuida et al., 1996, 1998; Yoshida et al., 1998), and the hematopoietic stem cell (HSC) compartment is expanded in the absence of caspase-3 (Janzen et al., 2008). Although this suggests an accumulation of cells otherwise destined to die, in both cases, the evidence points to a more complex mechanism. In the brain, controversy exists as to the extent of cell death in mice lacking the caspase cascade, and recent studies indicate that changes in morphogen gradients may underpin aberrant forebrain development (Honarpour et al., 2001; Nonomura et al., 2013; Oppenheim et al., 2001).

HSCs present a similar conundrum. HSC survival is governed by BCL-2 family proteins. Deletion of prosurvival Mcl-1 leads to their death, whereas overexpression of Bcl-2 increases their number (Domen et al., 2000; Opferman et al., 2005). This has led to a model whereby a proportion of HSCs undergo apoptosis during the normal course of hematopoiesis; hence, a reduction in apoptosis is proposed to lead to accumulation of HSCs in vivo (Orello and Dzierzak, 2007). The expansion of HSCs observed in caspase-3-deficient mice would accord with this notion. Intriguingly, however, the evidence suggests that, rather than accumulating through failure to die, *Casp3*^{-/-} HSCs proliferate due to abnormalities in cytokine signaling (Janzen et al., 2008), suggesting potential non-cell death roles for the apoptotic caspase cascade. In fact, apoptotic caspases are increasingly implicated in other cellular processes such as differentiation (Yi and Yuan, 2009). In some cases, these roles are a byproduct of, or are associated with, apoptosis; in others, they appear to be “nonapoptotic” in nature.

Here, we show that the caspase cascade functions during apoptosis to prevent dying cells from producing type I interferon (IFN). Bak- and Bax-mediated mitochondrial damage triggers the release of mitochondrial DNA (mtDNA), which is recognized by the cGAS/STING-mediated cytosolic DNA sensing pathway. In the absence of the apoptotic caspases, this leads to the induction of IFN- β transcription and IFN- β secretion by the dying cell. Loss of the caspase cascade leads to elevated IFN- β levels in vivo. This feeds back to, and has a profound impact on, the HSC compartment, which is highly sensitive to the effects of type I IFN. Thus the apoptotic caspase cascade regulates the immunological impact an apoptotic cell has on the host by preventing dam-

age-associated molecular pattern (DAMP) signaling induced by mtDNA.

RESULTS

HSC Expansion and Dysfunction in the Absence of Caspase-9

To define the requirement for the intrinsic apoptosis pathway (Figure 1A) in HSC homeostasis, we generated mice lacking Bak and Bax or caspase-9. Because these animals die postnatally, we first profiled the HSC-containing lineage⁻ Sca1⁺ Kit⁺ (LSK) population in fetal livers at embryonic (E) day 13.5. Relative to wild-type (WT) counterparts, the proportion of LSKs in *Bak*^{-/-} *Bax*^{-/-} fetal livers was unchanged (Figures 1B and 1C). In contrast, LSKs were increased ~5-fold in *Casp9*^{-/-} fetal livers. To establish whether this was hematopoietic cell intrinsic, we transplanted fetal liver cells (FLCs) into lethally irradiated WT recipients. 12–16 weeks posttransplantation, a small but statistically significant increase in LSKs was observed in mice that received *Bak*^{-/-} *Bax*^{-/-} FLCs (Figures 1D and 1E). Recapitulating the situation in the fetal liver, the bone marrow of mice reconstituted with *Casp9*^{-/-} cells contained 5-fold more LSKs than those transplanted with WT cells. Collectively, these data suggested that HSC numbers are expanded in the absence of caspase-9. This was a counterintuitive result, given that Bak and Bax are the critical mediators of the intrinsic apoptosis pathway, whereas the downstream caspase cascade is thought to be dispensable for cell death.

To examine HSC function, WT, *Bak*^{-/-} *Bax*^{-/-}, or *Casp9*^{-/-} fetal liver test cells were mixed 50:50 with WT competitor cells and transplanted into lethally irradiated WT recipients (Figure 1F). 16 weeks posttransplant, the contribution of *Bak*^{-/-} *Bax*^{-/-} cells to peripheral blood B and T lymphocytes, myeloid cells, and bone marrow LSKs significantly outweighed that of WT competitor cells (Figures 1G and S1A available online). In contrast, *Casp9*^{-/-} peripheral blood B and T lymphocytes, myeloid cells, and bone marrow LSKs were present in equal numbers to WT competitors, suggesting that, despite the aberrant LSK profile, the *Casp9*^{-/-} fetal liver contains either (1) normal numbers of HSCs or (2) more HSCs than WT, but they are functionally impaired. We therefore tested the self-renewal capacity of HSCs by harvesting bone marrow from primary recipients and transplanting it into secondary recipients (Figures 1H, 1I, and S1B). In contrast to WT and *Bak*^{-/-} *Bax*^{-/-} cells, *Casp9*^{-/-} bone marrow exhibited a profoundly reduced contribution to peripheral blood B and T lymphocytes, myeloid cells, and bone marrow LSKs at 16 weeks post-secondary transplantation (Figure 1I), indicating that HSC function is severely compromised in the absence of caspase-9.

HSC Dysfunction in Caspase-9-Deficient Mice Is Cell Extrinsic

To establish whether LSK expansion in *Casp9*^{-/-} mice was intrinsic to the LSK population itself, we generated mixed bone marrow chimeras by transplanting WT or *Casp9*^{-/-} E13.5 FLCs with WT filler FLCs into lethally irradiated recipients. Consistent with our previous observations, 12–16 weeks later, we observed an expansion of *Casp9*^{-/-} LSKs in the WT:*Casp9*^{-/-} chimeras.

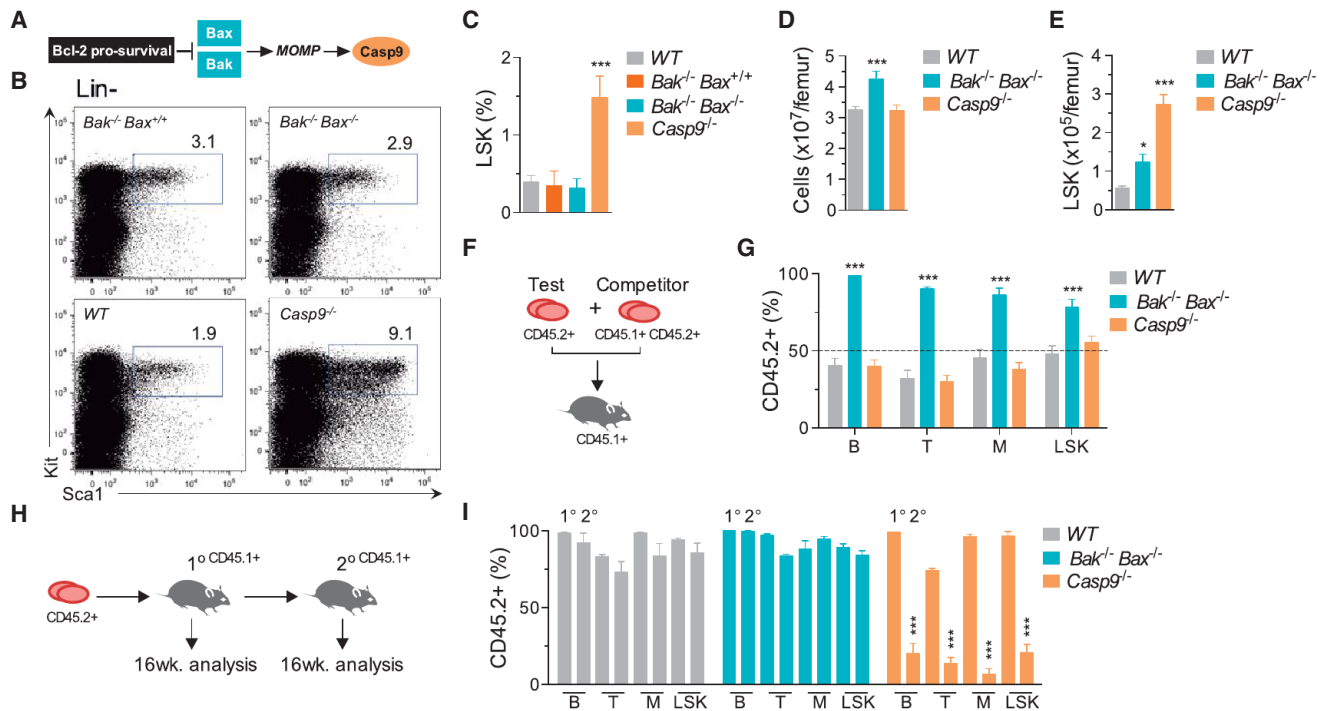


Figure 1. Deficiency of Caspase-9, but Not Bak and Bax, Impairs Hematopoietic Stem Cell Function

(A) The intrinsic apoptosis pathway.
 (B) Representative plots of LSK cell frequency in E13.5 fetal livers (gates display percentage of Lin⁻).
 (C) FACS analysis of LSKs in WT (n = 9), *Bak*^{-/-} *Bax*^{+/+} (n = 5), *Bak*^{-/-} *Bax*^{-/-} (n = 7), and *Casp9*^{-/-} (n = 10) E13.5 fetal livers.
 (D) Bone marrow cellularity in recipients of WT (n = 15), *Bak*^{-/-} *Bax*^{-/-} (n = 8), and *Casp9*^{-/-} (n = 9) FLCs 12–16 weeks posttransplant.
 (E) Donor-derived LSK cells in WT (n = 15), *Bak*^{-/-} *Bax*^{-/-} (n = 8), and *Casp9*^{-/-} (n = 9) bone marrow chimeras 12–16 weeks posttransplant.
 (F) Fetal liver competitive transplantation assay.
 (G) Proportion of CD45.2⁺ peripheral blood B lymphocytes, T lymphocytes, myeloid cells, and bone marrow LSK cells 16 weeks after competitive transplantation (n = 3–4 E13.5 test-CD45.2⁺ fetal livers per genotype and three recipients per donor mix). See also Figure S1.
 (H) Secondary transplantation assay.
 (I) Donor-CD45.2⁺ contribution to peripheral blood B lymphocyte, T lymphocyte and myeloid cells, and bone marrow LSK cells of 1° and 2° recipients 16 weeks posttransplant (n = 3–5 donor fetal livers per genotype and 2–3 recipients per donor bone marrow) See also Figure S1.
 Means were compared to WT using a one-way ANOVA with Bonferroni correction. Data represent the mean ± SEM. *p ≤ 0.05, **p ≤ 0.01, and ***p ≤ 0.005.

Strikingly, WT LSK numbers were also significantly increased relative to those in WT:WT chimeras (Figure 2A). Their Sca1 expression profile resembled that of *Casp9*^{-/-} Lin⁻ Kit⁺ cells (Figure 2B). Thus, WT LSKs expand in the presence of caspase-9-deficient hematopoietic cells, suggesting that HSC expansion in *Casp9*^{-/-} mice is driven by cell-extrinsic factors.

Caspase-9-Deficient HSPCs Exhibit a Type I Interferon Response Signature

We therefore analyzed the gene expression profile of Lineage⁻ Kit⁺ CD45.2⁺ hematopoietic stem and progenitor cells (HSPCs) harvested primary bone marrow chimeras. 495 genes (corresponding to 602 probes) were significantly upregulated in *Casp9*^{-/-} HSPCs (Figure 2C and Table S1 available online), with 275 genes (corresponding to 346 probes) differentially expressed (DE) in *Bak*^{-/-} *Bax*^{-/-} HSPCs. In *Casp9*^{-/-} HSPCs, gene set enrichment analysis using Camera identified enrichment for multiple type I IFN response signatures that were not apparent in *Bak*^{-/-} *Bax*^{-/-} HSPCs (Figures 2D and 2E). The top ten upregulated genes in *Casp9*^{-/-} HSPCs by fold change

were all type I IFN targets (Table S1). Quantitative PCR (qPCR) analysis of canonical type I IFN-stimulated genes (ISGs) confirmed their upregulation relative to WT in *Casp9*^{-/-} FLCs and bone marrow cells (Figures 2F and 2G). We therefore examined levels of the type I IFNs, IFNα and IFNβ, in the serum of primary transplant recipients. Whereas IFNα was undetectable in all genotypic classes, we observed significantly elevated IFNβ in mice reconstituted with *Casp9*^{-/-}, but not *Bak*^{-/-} *Bax*^{-/-}, cells (Figure 2H).

HSC Expansion in Caspase-9-Deficient Mice Is Caused by Type I Interferons

Type I IFNs have been shown to induce HSC proliferation, leading to functional exhaustion in vivo (Essers et al., 2009; Sato et al., 2009). We therefore generated mice doubly deficient for caspase-9 and the type I IFN receptor (Ifnar1). Although loss of Ifnar1 did not rescue postnatal lethality of *Casp9*^{-/-} mice (data not shown), deletion of Ifnar1 prevented LSK expansion in *Casp9*^{-/-} fetal livers (Figures 3A and 3B). Next, bone marrow chimeras were generated by reconstituting recipients

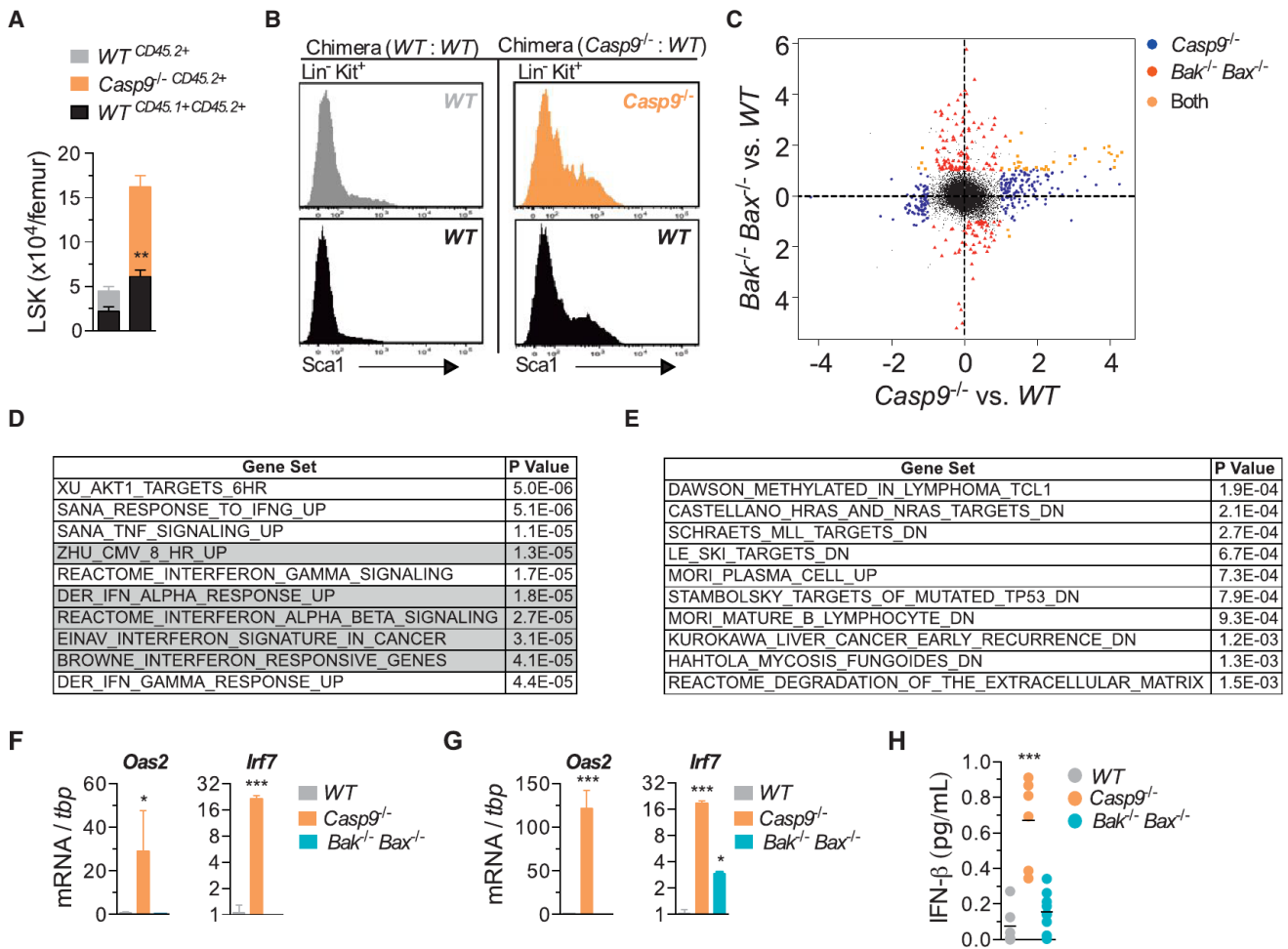


Figure 2. Loss of Caspase-9 Results in Elevated Type I Interferon In Vivo

(A) FACS analysis of mixed bone marrow chimeras 12–16 weeks posttransplant of WT (CD45.2⁺) or Casp9^{-/-} (CD45.2⁺) with WT “bystander” (CD45.1⁺CD45.2⁺) E13.5 FLCs into lethally irradiated CD45.1⁺ recipients. Number of donor-derived bone marrow LSK cells from both fractions is displayed (n = 8 mixed bone marrow chimeras per genotype from 3–4 fetal livers per genotype). Means of WT (CD45.1⁺CD45.2⁺) “bystander” cells were compared using a two-tailed t test.

(B) Sca1 expression on Lin⁻ Kit⁺ bone marrow cells from mixed bone marrow chimeras.

(C) Scatterplot of differentially expressed probes in microarray analysis of WT, Casp9^{-/-}, and Bak^{-/-} Bax^{-/-} Lin⁻ Kit⁺ CD45.2⁺ bone marrow cells. See also Table S1.

(D) Top ten gene sets (ranked by p value) from Gene Set Enrichment Analysis (GSEA) of Casp9^{-/-} in (C) (gray indicates type I IFN signatures).

(E) Top ten gene sets (ranked by p value) from GSEA of Bak^{-/-} Bax^{-/-} in (C).

(F and G) Real-time qPCR analysis of type I ISGs in fetal liver (F) and bone marrow cells (G) (n = 3–4 E13.5 fetal livers and 3–4 bone marrow chimeras per genotype).

(H) IFN-β protein in serum of WT (n = 6), Casp9^{-/-} (n = 6), and Bak^{-/-} Bax^{-/-} (n = 8) bone marrow chimeras.

Unless indicated, means were compared to WT using a one-way ANOVA with Bonferroni correction. Data represent the mean ± SEM. *p ≤ 0.05, **p ≤ 0.01, and ***p ≤ 0.005.

with *Ifnar1*^{-/-} Casp9^{+/+}, *Ifnar1*^{+/+} Casp9^{-/-}, or *Ifnar1*^{-/-} Casp9^{-/-} FLCs. 16 weeks posttransplantation, recipients of *Ifnar1*^{+/+} Casp9^{-/-} cells exhibited a 5-fold increase in LSKs relative to mice that received *Ifnar1*^{-/-} Casp9^{+/+} cells (Figures 3C and 3D). In contrast, LSK numbers in *Ifnar1*^{-/-} Casp9^{-/-} chimeras were normal. Furthermore, deletion of *Ifnar1* restored the ability of Casp9^{-/-} HSCs to engraft the host and contribute to all major lineages upon secondary transplantation (Figures 3E and 3F). Collectively, our data indicate that loss of caspase-9 in vivo leads to production of IFN-β, which feeds back to the HSC compartment, resulting in loss of self-renewal capacity.

Apoptotic Mouse and Human Cells Produce Type I IFN When Caspases Are Inhibited

Activation of the apoptotic caspase cascade is impaired in the absence of caspase-9 (Li et al., 1997). We hypothesized that hematopoietic cells undergoing caspase-inhibited apoptosis might be the source of IFN-β. To test this, we treated WT murine splenocytes with the proapoptotic BH3 mimetic drug ABT-737, which targets the prosurvival proteins Bcl-x_L and Bcl-2 (Oltersdorf et al., 2005). ABT-737 induced caspase activation and cell death in splenocytes (Figures 4A and 4B). No IFN-β was detected in culture media of cells treated with ABT-737 alone. In

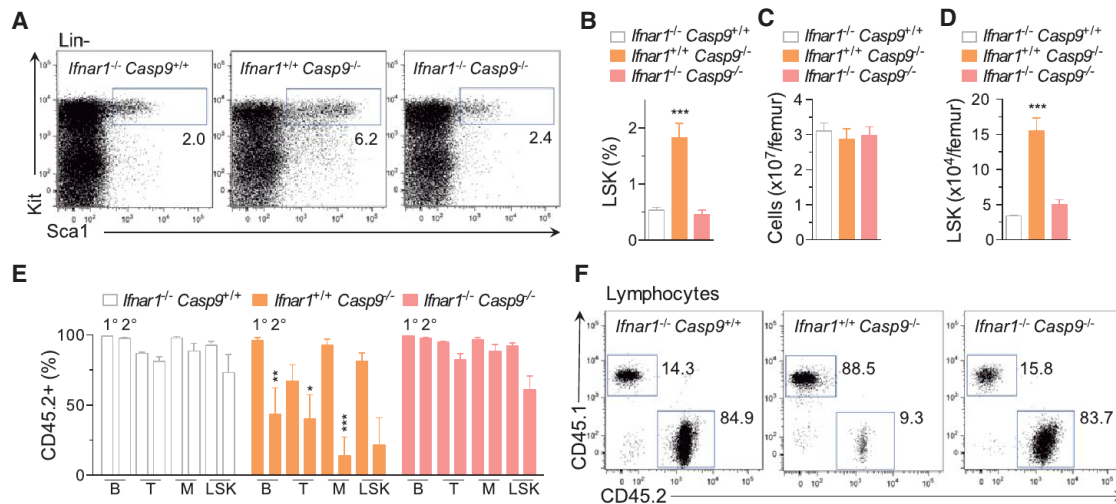


Figure 3. Type I Interferon Mediates the Hematopoietic Stem Cell Dysfunction Associated with Caspase-9 Loss

(A) Representative plots of LSK cell frequency in E13.5 fetal livers (gates display percentage of Lin⁻). (B) FACS analysis of LSKs in *Ifnar1^{-/-} Casp9^{+/+}* (n = 6), *Ifnar1^{+/+} Casp9^{-/-}* (n = 4), and *Ifnar1^{-/-} Casp9^{-/-}* (n = 6) E13.5 fetal liver. (C) Bone marrow cellularity from *Ifnar1^{-/-} Casp9^{+/+}* (n = 8), *Ifnar1^{+/+} Casp9^{-/-}* (n = 4), and *Ifnar1^{-/-} Casp9^{-/-}* (n = 5) bone marrow chimeras, 16 weeks posttransplant. (D) Number of donor-derived LSK cells from *Ifnar1^{-/-} Casp9^{+/+}* (n = 7), *Ifnar1^{+/+} Casp9^{-/-}* (n = 4), and *Ifnar1^{-/-} Casp9^{-/-}* (n = 5) bone marrow chimeras, 16 weeks posttransplant. (E) Donor-CD45.2⁺ contribution to the peripheral blood B lymphocyte, T lymphocyte, myeloid cells, and bone marrow LSK cells of 1° and 2° recipients at 16 weeks posttransplant. *Ifnar1^{-/-} Casp9^{+/+}* (n = 4), *Ifnar1^{+/+} Casp9^{-/-}* (n = 3), and *Ifnar1^{-/-} Casp9^{-/-}* (n = 4) donor fetal livers per genotype and three recipients per donor bone marrow. (F) Plots of representative analysis of donor contribution to 2° recipient lymphoid lineages in (E). Means were compared to WT using a one-way ANOVA with Bonferroni correction. Data represent the mean ± SEM. *p ≤ 0.05, **p ≤ 0.01, and ***p ≤ 0.005.

contrast, when apoptosis was triggered in the presence of the pan-caspase inhibitor Q-VD-Oph, IFN-β was produced (Figure 4C). To test whether this mechanism is conserved between mice and humans, peripheral blood mononuclear cells (PBMCs) were isolated from the blood of five healthy adult donors and treated with ABT-737. Caspase-3/7 activation and loss of cell viability was observed over 24 hr (Figures 4D and 4E). Upon co-incubation with ABT-737 and Q-VD-Oph, IFN-β secretion was observed in all five human PBMC samples (Figure 4F). These data demonstrated that, in both human and murine hematopoietic cells, caspase-inhibited apoptosis results in the production of IFN-β.

Apoptotic MEFs Produce Type I IFN When Caspases Are Inhibited

To better examine the role of the intrinsic apoptosis pathway, we utilized immortalized mouse embryonic fibroblasts (MEFs). WT MEFs are dependent on the prosurvival proteins Mcl-1 and Bcl-x_L for survival (Figure 5A). MEFs lacking Mcl-1 undergo Bak- and Bax-mediated apoptosis in response to ABT-737 (van Delft et al., 2006). We therefore treated multiple *Mcl1^{-/-}* MEF lines with increasing concentrations of ABT-737 and co-incubated them with either Q-VD-Oph or another pancaspase inhibitor, z-VAD.fmk. ABT-737 treatment of *Mcl1^{-/-}* MEFs induced caspase-3/7 activity and cell death (Figures 5B and 5C). Coincubation with either z-VAD.fmk or Q-VD-Oph blocked caspase activation and prevented loss of viability. Although undetectable in supernatant from *Mcl1^{-/-}* MEFs treated with

ABT-737 or caspase inhibitor alone, IFN-β was induced when ABT-737 (or other apoptotic stimuli) (Figure S2) was combined with z-VAD.fmk or Q-VD-Oph (Figure 5D). Upregulation of *Ifnb1*, the gene encoding IFN-β, was evident 4 hr posttreatment (Figure 5E). These data demonstrate that nonhematopoietic cells also produce IFN-β when undergoing caspase-inhibited apoptosis and implicate Bak and Bax as the initiators of the signal that triggers IFN-β production.

The Apoptotic Caspase Cascade Suppresses Bak/Bax-Mediated Type I IFN Production

To confirm whether Bak and Bax activation induces IFN-β production, we triggered mitochondrial apoptosis in WT, *Casp9^{-/-}*, and *Bak^{-/-} Bax^{-/-} Casp9^{-/-}* MEF lines by combining ABT-737 treatment with expression of Bim_{S2A}, which targets Mcl-1 (Lee et al., 2008). Expression of Bim_{S4E}, an inert form of Bim, served as a negative control. *Casp3^{-/-} Casp7^{-/-}* MEF lines were included to determine whether deletion of the effector caspases causes IFN production. In WT cells expressing Bim_{S2A}, ABT-737 treatment induced a 3-fold increase in caspase activity and an ~75% loss of viability (Figures 5F–5H). In contrast, both *Casp9^{-/-}* and *Bak^{-/-} Bax^{-/-} Casp9^{-/-}* cells were resistant to Bim_{S2A} + ABT-737. Analysis of the supernatant demonstrated abundant IFN-β secretion by *Casp9^{-/-}* and *Casp3^{-/-} Casp7^{-/-}* cells treated with Bim_{S2A} + ABT-737, but not WT or *Bak^{-/-} Bax^{-/-} Casp9^{-/-}* counterparts (Figures 5I and S3). These data demonstrated that, in vitro, induction of Bak- and Bax-mediated apoptosis stimulates type I IFN

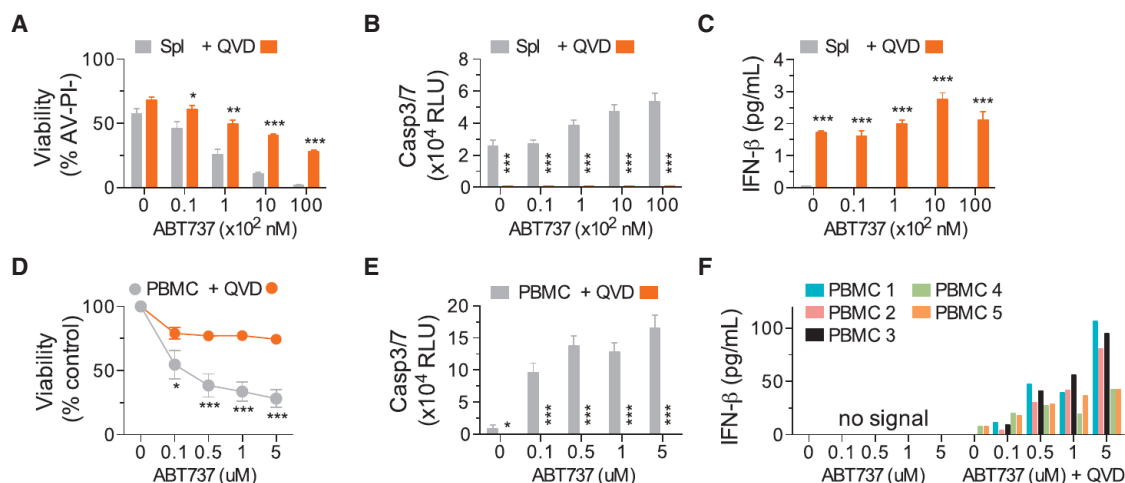


Figure 4. Apoptotic Hematopoietic Cells Produce Type I Interferon When Caspases Are Inhibited

(A) Viability of murine splenocytes treated with ABT-737 \pm 20–30 μ M Q-VD-OPH (QVD) for 24 hr. (B and C) (B) Caspase activity and (C) IFN- β in culture supernatant after 24 hr of treatment ($n = 4$ mice). Means were compared using a two-tailed t test. Data represent the mean \pm SEM. * $p \leq 0.05$, ** $p \leq 0.01$, and *** $p \leq 0.005$. (D) Percentage of viable human PBMCs as quantitated by ATP levels (CellTiterGlo) after 24 hr of treatment with ABT-737 and coincubation with 20–30 μ M Q-VD-OPH (QVD). (E) Bar graphs of caspase activity after 6 hr and (F) IFN- β in culture supernatant after 24 hr ($n = 5$ healthy donor blood samples). Representative of two independent experiments.

production when either the apoptotic initiator (caspase-9) or effector (caspase-3/7) caspases are inactivated.

We reasoned that eliminating Bak and Bax would remove the stimulus for IFN- β production in caspase-deficient mice. To test this, we generated *Casp9*^{−/−}, *Apaf1*^{−/−}, *Casp3*^{−/−}, *Casp3*^{−/−} *Casp7*^{−/−}, and *Bak*^{−/−} *Bax*^{−/−} *Casp9*^{−/−} bone marrow chimeras. 16 weeks posttransplantation, LSKs were increased in recipients of *Apaf1*^{−/−}, *Casp9*^{−/−} or *Casp3*^{−/−} *Casp7*^{−/−} cells, demonstrating that inhibition of the apoptotic caspase cascade at either the apoptosome or effector stage triggers LSK expansion (Figures 5J and 5K). In contrast, the LSK profile in *Bak*^{−/−} *Bax*^{−/−} *Casp9*^{−/−} chimeras was unchanged. Thus, in the absence of Bak- and Bax-mediated apoptosis, deletion of caspases does not cause LSK expansion. Accordingly, IFN- β levels in *Bak*^{−/−} *Bax*^{−/−} *Casp9*^{−/−} chimeras were equivalent to WT, and no upregulation of type I IFN response genes was apparent (Figures 5L and 5M). Secondary bone marrow transplants confirmed that loss of Bak/Bax-mediated apoptosis also rescued HSC function, with *Bak*^{−/−} *Bax*^{−/−} *Casp9*^{−/−} bone marrow efficiently engrafting the host and contributing to the LSK compartment and all major lineages (Figure 5N). Thus, deletion of Bak and Bax prevents IFN- β production and HSC dysfunction in mice lacking a functional apoptotic caspase cascade.

Dying Cells Upregulate Type I IFN Production in a Cell-Intrinsic Manner

We next examined whether apoptotic cells secrete IFN- β in a cell-intrinsic manner. *Ifnar1*^{−/−} *Casp9*^{−/−} MEFs were transduced with Bim_s2A-GFP or Bim_s4E-GFP and co-plated with unmanipulated *Ifnar1*^{−/−} MEFs (Figures 6A and 6B). *Ifnar1*-deficient cells were utilized to eliminate paracrine/autocrine effects of IFN- β . Cocultures were treated with ABT-737, which induced

apoptosis in the *Ifnar1*^{−/−} *Casp9*^{−/−} Bim_s2A-GFP cells, but not the other three cell populations. Subsequently, GFP⁺ (apoptotic or nonapoptotic) and GFP[−] (bystander) cells were sorted, and expression of *Ifnb1* was analyzed by qPCR. Relative to healthy bystanders, a significant upregulation of *Ifnb1* mRNA was observed in apoptotic, but not nonapoptotic, cells (Figures 6B and 6C). When using WT (rather than *Ifnar1*^{−/−}) bystanders, type I IFN response genes were strongly induced in bystanders cocultured with apoptotic cells (Figure 6D), indicating that cells undergoing caspase-inhibited apoptosis actively transcribe *Ifnb1* and secrete bioactive IFN- β .

Apoptotic IFN Production Is Driven by Mitochondrial DNA

Considering the mechanism by which Bak and Bax stimulate a cell to produce type I IFN, we reasoned that release of a mitochondrial factor into the cytoplasm could be the initiating event. In the context of microbial invasion, type I IFNs are induced by a range of pathogen-associated molecular patterns (PAMPs). Viral nucleic acids are an important, and potent, example. Their presence is detected by cytosolic receptors that activate signaling cascades leading to the upregulation of IFN transcription (Paludan and Bowie, 2013). We hypothesized that Bak- and Bax-mediated damage to mitochondria may cause the release of mtDNA and that the latter would act as a DAMP capable of recognition by the dying cell's innate nucleic acid sensors. We therefore generated mtDNA-depleted cells (so-called “ ρ^0 ” cells) by culturing *Mcl1*^{−/−} MEFs in ethidium bromide (King and Attardi, 1989). qPCR analysis revealed a near-complete absence of mtDNA after three passages (Figures 6E and S4). Expression profiling of ρ^0 cells demonstrated that 28 genes (corresponding to 42 probes)

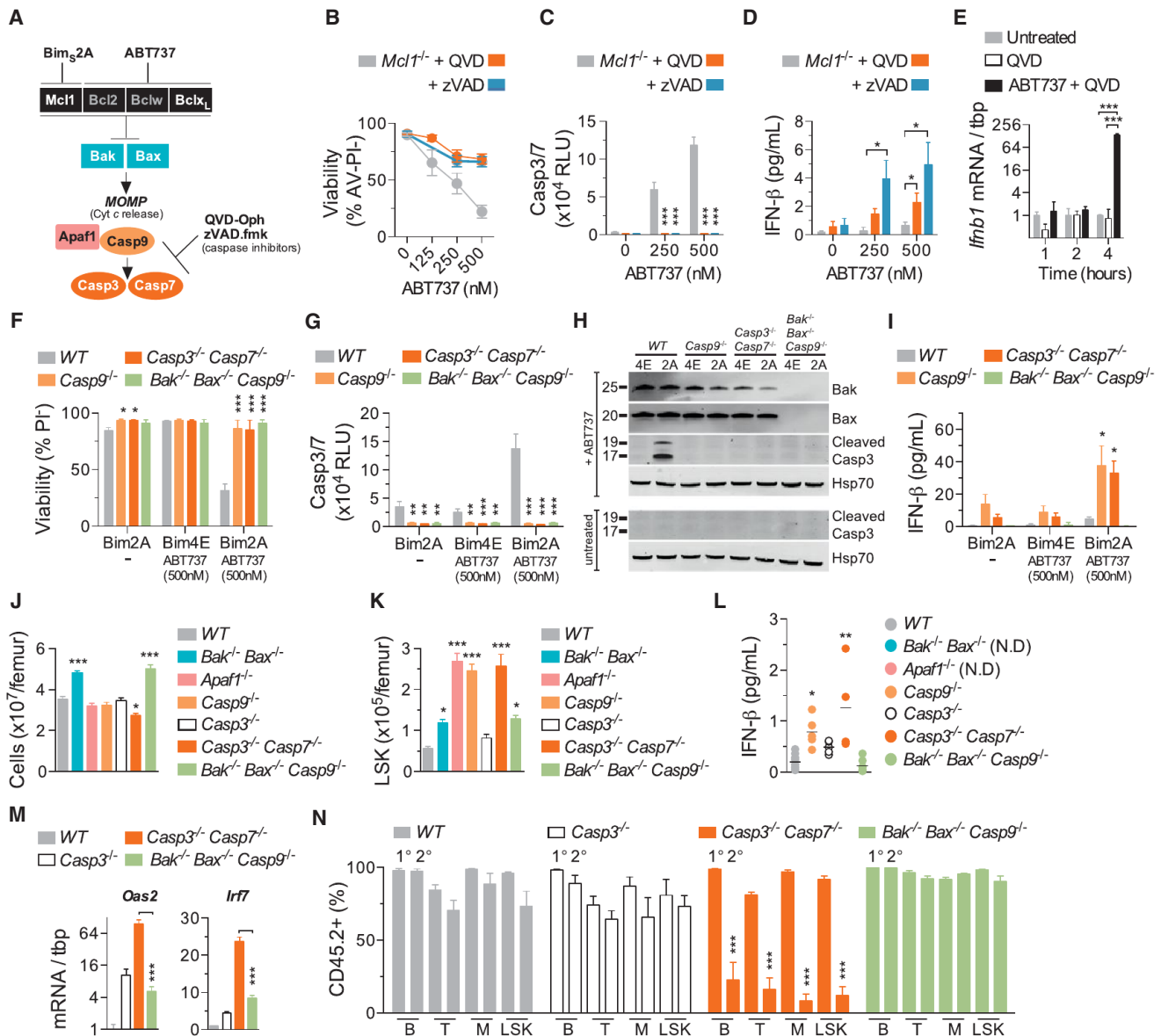


Figure 5. Disabling Caspases Downstream of Bak and Bax Triggers Type I IFN Production

(A) Schematic diagram of the manipulation of intrinsic apoptosis in MEFs.

(B–D) (B) Viability of MEFs treated with ABT-737 ± 20–30 μM of Q-VD-Oph (QVD) or z-VAD.fmk (zVAD) for 24 hr, (C) bar graphs of caspase activity after 6 hr, and (D) IFN- β in supernatant after 24 hr (n = 9 independent MEF lines). Means were compared using a two-tailed t test. See also Figure S2.

(E) Real-time qPCR analysis of *Ifnb1* induction in *Mcl1*^{-/-} MEFs (n = 3 independent MEF lines). Means were compared using a two-tailed t test.

(F and G) (F) Viability of MEFs expressing Bim2A or Bim4E and treated with ABT-737 for 20–24 hr and (G) caspase activity after 6 hr.

(H and I) (H) Immunoblot of lysates of MEFs treated with ABT-737 (1 μM) for 4 hr and (I) bar graph of IFN- β in supernatant after 20–24 hr (n = 3 independent MEF lines per genotype). See also Figures S2 and S3.

(J) Bone marrow cellularity from WT (n = 22), *Bak*^{-/-} *Bax*^{-/-} (n = 8), *Apaf1*^{-/-} (n = 5), *Casp9*^{-/-} (n = 11), *Casp3*^{-/-} (n = 13), *Casp3*^{-/-} *Casp7*^{-/-} (n = 7), and *Bak*^{-/-} *Bax*^{-/-} *Casp9*^{-/-} (n = 9) bone marrow chimeras.

(K) Number of donor-derived LSK cells from WT (n = 22), *Bak*^{-/-} *Bax*^{-/-} (n = 8), *Apaf1*^{-/-} (n = 5), *Casp9*^{-/-} (n = 11), *Casp3*^{-/-} (n = 13), *Casp3*^{-/-} *Casp7*^{-/-} (n = 7), and *Bak*^{-/-} *Bax*^{-/-} *Casp9*^{-/-} (n = 9) bone marrow chimeras.

(L) IFN- β in serum of WT (n = 10), *Casp9*^{-/-} (n = 5), *Casp3*^{-/-} (n = 5), *Casp3*^{-/-} *Casp7*^{-/-} (n = 4), and *Bak*^{-/-} *Bax*^{-/-} *Casp9*^{-/-} (n = 7) bone marrow chimeras. Not done, N.D.

(M) Real-time qPCR analysis of type I ISGs in bone marrow cells (n = 3–4 bone marrow chimera per genotype). Means were compared using a two-tailed t test.

(N) Donor-CD45.2⁺ contribution to the peripheral blood B lymphocyte, T lymphocyte, myeloid cells, and bone marrow LSK cells of 1° and 2° recipients 16 weeks posttransplant (n = 3 donor fetal livers per genotype and 3 recipients per donor bone marrow).

Unless otherwise indicated, means were compared to WT using a one-way ANOVA with Bonferroni correction. Data represent the mean ± SEM. *p ≤ 0.05, **p ≤ 0.01, and ***p ≤ 0.005.

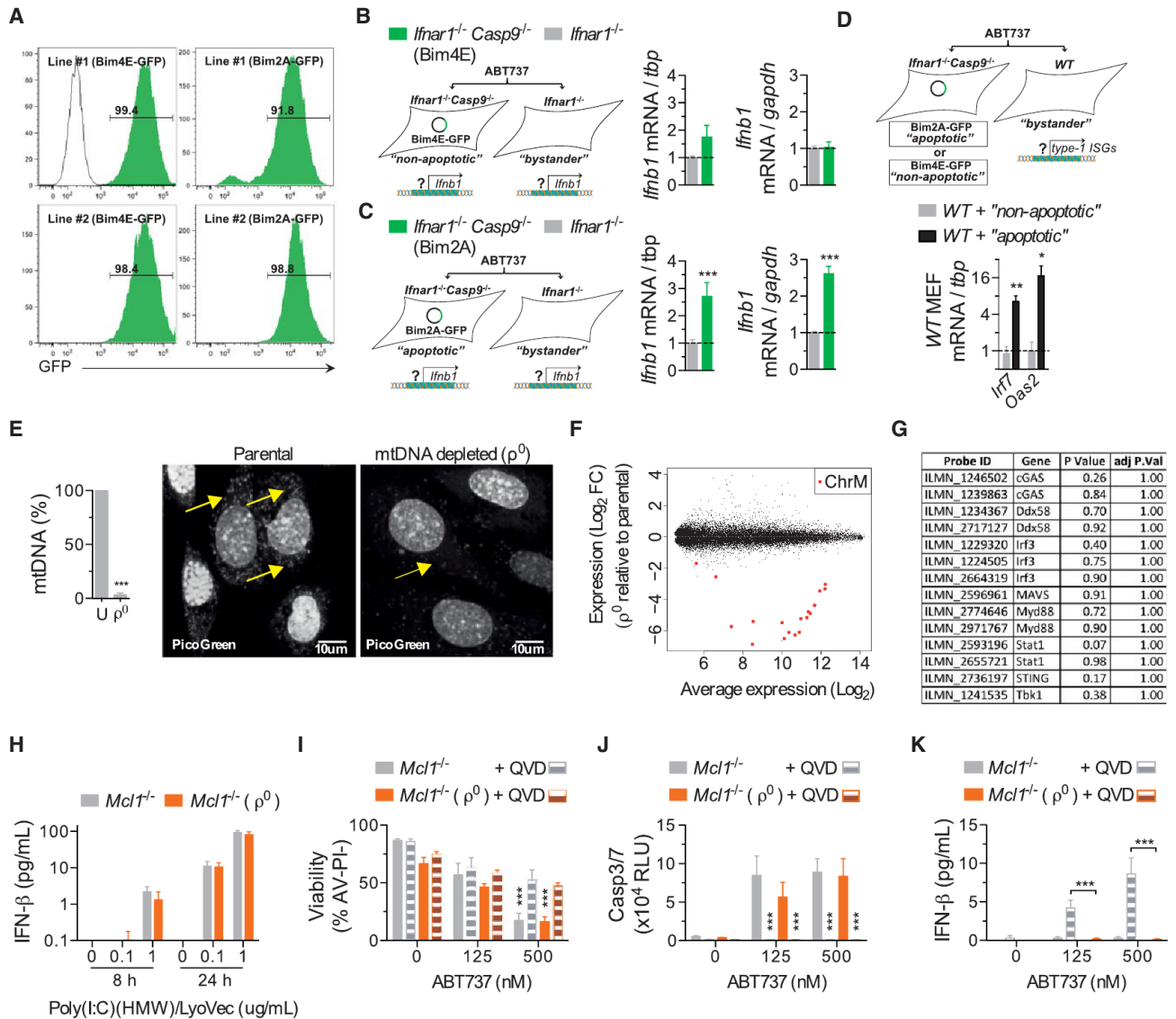


Figure 6. mtDNA Triggers Type I IFN Production during Caspase-Inhibited Apoptosis

(A) Representative plots of infection efficiency of the MEFs used in (B–D).
 (B and C) Real-time qPCR analysis of nonapoptotic (expressing Bim₄E, B) and apoptotic (expressing Bim₂A, C) MEFs and their respective *Ifnar1*^{−/−} bystander from cocultures after treatment with ABT-737 (500 nM) for 18–20 hr (data combined from three experiments, with two independent MEF lines per genotype). mRNA expression is shown relative to two independent housekeeping genes, *tbp* and *gapdh*.
 (D) Real-time qPCR analysis of WT bystanders from cocultures after treatment with ABT-737 for 18–20 hr (data combined from two experiments, with two independent MEF lines/genotype).
 (E) Real-time qPCR analysis of mtDNA content from *Mcl1*^{−/−} MEFs cultured in ethidium bromide to generate mtDNA-depleted (ρ^0) MEFs (used in F–J). Representative image of MEFs stained with PicoGreen nucleic acid stain. Arrows indicate mtDNA. Scale bars, 10 μ m. See also Figure S4.
 (F) Scatterplot of differentially expressed probes from microarray analysis of *Mcl1*^{−/−} ρ^0 MEFs compared to their respective parental *Mcl1*^{−/−} MEF. (n = 3 independent MEF lines). Chromosome M, ChrM. See also Table S2.
 (G) Table of a selected set of type I IFN response genes from analysis in (F).
 (H) Bar graph of IFN- β in the supernatant of ρ^0 and parental MEFs transfected with Poly(I:C)(HMW) (n = 3 independent MEF lines).
 (I–K) (I) Bar graphs of the viability of ρ^0 and parental MEFs treated with ABT-737 \pm 20–300 μ M of Q-VD-Oph (QVD) for 24 hr, (J) caspase activity after 6 hr, and (K) IFN- β in supernatant after 24 hr (n = 4 independent MEF lines).
 Means were compared using a two-tailed t test. Data represent the mean \pm SEM. *p \leq 0.05, **p \leq 0.01, and ***p \leq 0.005.

were downregulated and 61 genes (corresponding to 82 probes) were upregulated (Figure 6F and Table S2). 18 probes corresponding to 9 mitochondrially encoded genes were pre-

sent on the array. These were the most downregulated genes in ρ^0 cells (Camera p value = 0.0095). Gene set analysis using Camera detected no significant enrichment for any of the c2

expression signatures (data not shown). In addition, a subset of known type I IFN response genes was not significantly downregulated in ρ^0 cells (Figure 6G, Camera p value for downregulation = 0.45). Consistent with there being no functional impairment of IFN response pathways, ρ^0 cells responded normally when transfected with poly(I:C) (Figure 6H). Treatment of *Mcl1*^{-/-} ρ^0 cells with ABT-737 resulted in caspase activation and a loss of viability similar in magnitude to that exhibited by the parental lines (Figures 6I and 6J). In the presence of Q-VD-Oph, IFN- β production was observed in the parental, but not ρ^0 , cells (Figure 6K). These data indicate that mtDNA is the trigger for IFN production downstream of Bak- and Bax-mediated mitochondrial damage.

Bak/Bax-Mediated mtDNA Release Triggers cGAS/STING-Dependent IFN Production

Two major pathways that mediate type I IFN production in response to intracellular microbial DNA have been described (Paludan and Bowie, 2013). The first is initiated by Toll-like receptor 9 (Tlr9) recognition of DNA localized to endosomes, which triggers Myd88 signaling. Deletion of Myd88 in *Casp9*^{-/-} chimeras failed to normalize serum IFN- β levels and did not prevent LSK expansion (Figure S5), suggesting that mtDNA-induced IFN- β production does not require the Tlr-mediated endosomal recognition pathway. The second is the stimulator of interferon genes (STING) pathway (Barber, 2014). It is triggered when DNA binds to cyclic GMP-AMP synthase (cGAS), thereby catalyzing production of cyclic GMP-AMP dinucleotide (cGAMP), which binds to and activates STING. Activated STING induces type I IFN transcription via the Tbk1-Irf3 signaling axis. We treated splenocytes from WT and STING-deficient mice with ABT-737 in the presence or absence of Q-VD-Oph. ABT-737 induced apoptotic caspase activation and cell death in WT and *STING*^{-/-} cells with similar kinetics (Figures 7A and 7B). Coincubation with Q-VD-Oph triggered IFN- β secretion by WT, but not *STING*^{-/-}, cells (Figure 7C).

To further dissect the requirement for the STING pathway, we derived MEFs from *STING*^{-/-} mice and generated MEF lines lacking cGAS and Irf3 by CRISPR/Cas9-mediated gene targeting. All lines underwent apoptosis in response to ABT-737 and Bim_S2A (Figures 7D and 7E). WT cells produced IFN- β when coincubated with Q-VD-Oph. In contrast, *cGAS*^{CRISPR/-}, *STING*^{-/-}, and *Irf3*^{CRISPR/-} cells did not (Figure 7F). The requirement for Tbk1 was examined by pretreating *Mcl1*-deficient MEFs with the Tbk1 inhibitor MRT-67307. *Mcl1*^{-/-} cells undergoing caspase-inhibited apoptosis secreted IFN- β (Figures 7G and 7H). Addition of MRT-67307 suppressed IFN- β production by dying cells (Figure 7I). These data demonstrate that caspase-inhibited apoptosis triggers cGAS/STING/Tbk1/Irf3-mediated IFN- β production.

Recent evidence indicates that cGAS is activated via a direct interaction with cytosolic DNA (Civril et al., 2013; Sun et al., 2013). To establish whether mtDNA interacts with cGAS during caspase-inhibited apoptosis, we immunoprecipitated cGAS from *Mcl1*^{-/-} MEFs treated with ABT-737 and Q-VD-Oph and utilized qPCR analysis to detect coprecipitated DNA. Although there was no mtDNA enrichment when cGAS was immunoprecipitated from untreated cells, a significant enrichment for

mtDNA, but not genomic DNA, was observed in cells treated with ABT-737 and Q-VD-Oph (Figure 7J). Collectively, these data indicate that mtDNA is released, binds to, and activates cGAS during caspase-inhibited apoptosis.

DISCUSSION

Role of the Intrinsic Apoptosis Pathway in HSC Homeostasis

Deletion of Bak and Bax had no impact on the number of immunophenotypic HSCs in the fetal liver, suggesting that the intrinsic apoptosis pathway is dispensable for HSC homeostasis during development. Upon transplantation, recipients of *Bak*^{-/-} *Bax*^{-/-} FLCs exhibited a statistically significant 2-fold increase in bone marrow LSK number. This increase is similar to that reported for mice overexpressing Bcl-2 (Domen et al., 2000) and supports a role for the intrinsic pathway in regulating adult HSC homeostasis. However, a simpler explanation might be that HSCs lacking Bak and Bax are more resistant to the stresses of transplantation. Consistent with this, *Bak*^{-/-} *Bax*^{-/-} FLCs outcompeted WT counterparts in mixed transplants. Similar effects have been reported for HSCs lacking Bim (Labi et al., 2013). Thus, the extent to which death via the intrinsic apoptosis pathway shapes the HSC pool at steady state remains to be determined. It may be that programmed cell death does not represent a significant fate for HSCs. Alternatively, other cell death modalities such as the extrinsic apoptosis pathway or necroptosis might contribute to HSC homeostasis.

Apoptotic versus Nonapoptotic Roles for Apoptotic Caspases

Apoptotic caspases have been ascribed a number of functions beyond the internal demolition of dying cells. Some of these (e.g., prostaglandin-induced tumor cell repopulation [Huang et al., 2011] and AMPA receptor internalization [Li et al., 2010b]) appear to be a byproduct of apoptotic cell death. Others (e.g., iPS cell reprogramming [Li et al., 2010a] and microglia activation [Burguillos et al., 2011]) are thought to represent “nonapoptotic” roles for caspases. Our genetic experiments, both in vitro and in vivo, demonstrate that, in suppressing IFN production, the apoptotic caspases play an apoptotic role, i.e., function downstream of Bak and Bax. Unless Bak and Bax are activated, mtDNA is not released, and caspases are not required to attenuate mtDNA-induced DAMP signaling. This has two important implications. First, blocking the apoptotic caspase cascade during apoptosis triggers the production of IFN- β , a potentially significant confounding experimental factor. This should be considered whenever the apoptotic caspase cascade is genetically or pharmacologically manipulated. Perturbations in type I IFN signaling may explain some of the published biological roles for caspases. Second, the fact that caspase-deficient mice present with a dramatic increase in LSKs, whereas *Bak*^{-/-} *Bax*^{-/-} animals do not, initially suggested that caspases play a “nonapoptotic” role in HSCs. However, the subsequent experiments with *Bak*^{-/-} *Bax*^{-/-} *Casp9*^{-/-} cells and bone marrow chimeras demonstrated that the phenotype is dependent on Bak/Bax activation. This highlights the importance of establishing whether apoptotic caspase activation in a given setting is the result of

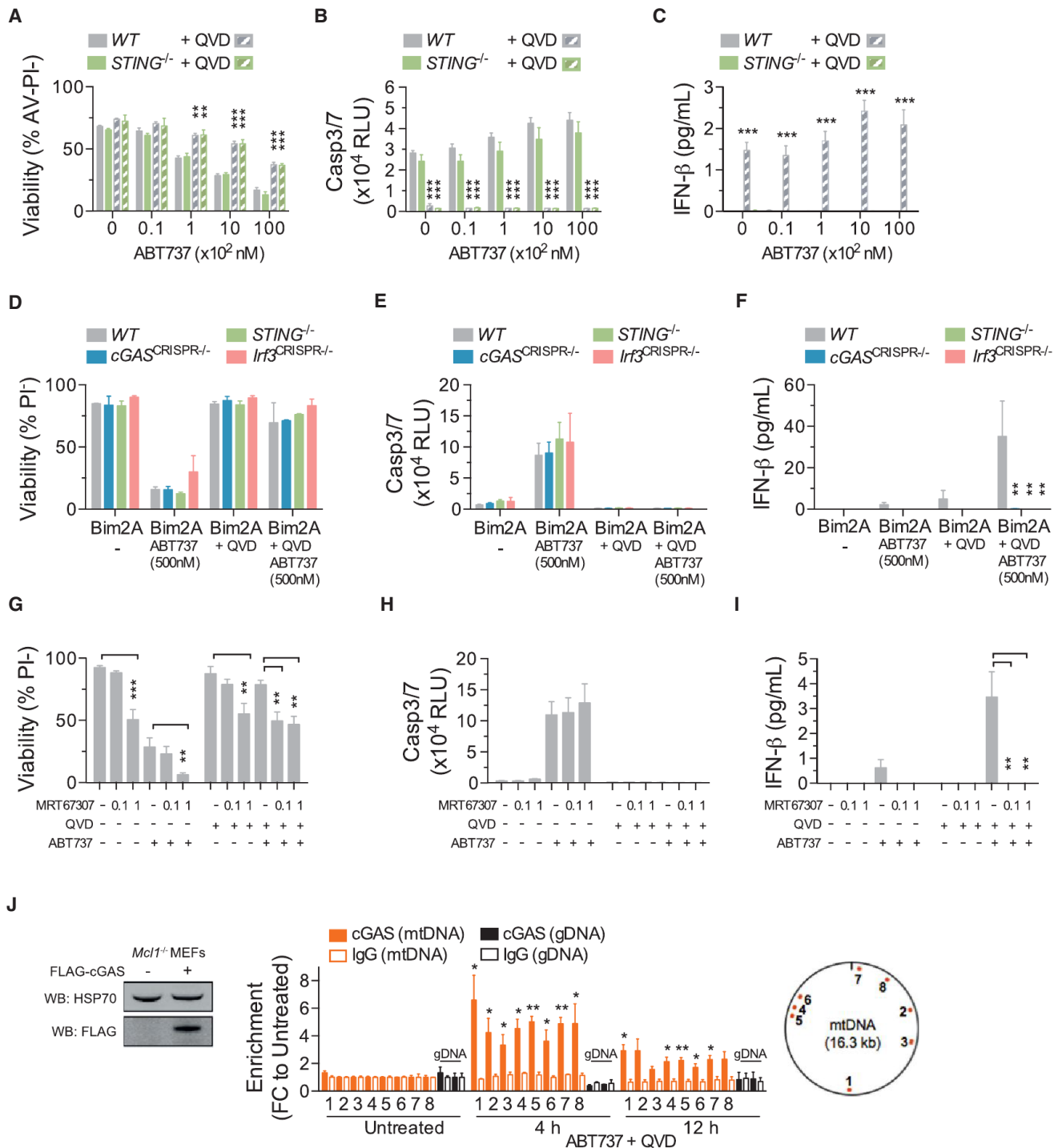


Figure 7. mtDNA Release into the Cytosol Triggers cGAS-STING-Tbk1-Irf3-Mediated Type I Interferon Production

(A–C) (A) Viability of murine splenocytes treated with ABT-737 \pm 20–30 μ M Q-VD-Oph (QVD) for 24 hr, (B) caspase activity after 6 hr, and (C) IFN- β in culture supernatant after 24 hr ($n = 4$ mice per genotype). Means were compared to WT using a one-way ANOVA with Bonferroni correction.

(D–F) (D) Viability of MEFs treated with ABT-737 \pm 20–30 μ M Q-VD-Oph (QVD) for 24 hr, (E) caspase activity after 6 hr, and (F) IFN- β in culture supernatant after 24 hr ($n = 3$ independent MEF lines per genotype, or 3 independent CRISPR/Cas9-targeted MEF clones). Means were compared to WT using a one-way ANOVA with Bonferroni correction.

(G–I) (G) Bar graphs of the viability of *Mcl1*^{-/-} MEFs pretreated for 1 hr with MRT-67307 (μ M) followed by ABT-737 (500 nM) \pm 20–30 μ M Q-VD-Oph (QVD) for 24 hr, (H) caspase activity after 6 hr, and (I) IFN- β in culture supernatant after 24 hr ($n = 3$ independent MEF lines).

(legend continued on next page)

upstream mitochondrial damage or an alternative signaling mechanism.

mtDNA Activates the STING-Dependent Cytosolic DNA Sensing Pathway

Our data indicate that Bak/Bax-dependent mtDNA release triggers the cGAS/STING-dependent cytosolic DNA sensing pathway. mtDNA released during cell death has been previously reported to provide a second signal that cooperates with signal 1 (e.g., LPS) to activate the NLRP3 inflammasome and induce IL-1 β production (Shimada et al., 2012). Whether mtDNA released via Bak/Bax also activates the inflammasome is unclear. However, other than IFN- β , mice with a caspase-9-deficient hematopoietic system did not exhibit elevated proinflammatory cytokine serum levels or any signs of systemic autoinflammatory disease when aged to 12 months (Figure S6). This suggests that, at least in the context of steady-state hematopoietic cell turnover, Bak/Bax-mediated mtDNA release does not result in unbridled inflammasome activity when the caspase cascade is inactivated. Whether the elevations in IFN- β caused by loss of the caspase cascade lead to the development of autoimmune disease over the longer term and whether perturbations in caspase activity might contribute to human autoinflammatory/autoimmune disease remain to be established.

Mechanisms of Caspase-Mediated Suppression of Type I IFN

During apoptosis, caspases orchestrate a global program of cellular demolition, targeting ~1,000 protein substrates (Crawford and Wells, 2011). They cause DNA damage, suppress transcription, shut down protein translation, and disable a host of other essential cellular processes. Unlike genomic damage, which is triggered by cleavage of iCAD, the inhibitor of caspase-activated DNase (CAD) (Enari et al., 1998), and PS exposure, which is facilitated by cleavage of Xkr8 (Suzuki et al., 2013), it seems likely that caspase-mediated suppression of IFN- β production (and DAMP signaling generally) is the result of multiple redundant processes. First, the fact that apoptotic CAD^{CRISPR-/-} MEFs secrete detectable amounts of IFN- β suggests that CAD-mediated genomic damage contributes to the attenuation of gene expression (Figure S7). Second, caspase-3/7 might cleave and inactivate a component or components of the type I IFN production pathway. There is evidence that IFN signaling intermediates, including Irf3, can be targeted by caspases (Crawford et al., 2013). Third, caspase-3/7 could mediate the degradation of mtDNA, thereby preventing its interaction with cGAS. This might occur via CAD or perhaps lysosomal deoxyribonuclease (DNase) II, which can digest the genomic DNA of engulfed cells. Although the precise contribution these and other mechanisms make to suppressing IFN- β

production in apoptotic cells remains to be established, it seems unlikely that any one of them is solely responsible. We suggest that caspase-mediated acceleration of cell death and suppression of DAMP signaling are inextricably linked. Both outcomes are achieved via global cellular demolition.

Potential Crosstalk between Apoptosis and Necroptosis

Necroptosis is a regulated form of necrotic cell death mediated by receptor-interacting protein kinase 3 (RIPK3) and the pseudokinase M1K1 (Linkermann and Green, 2014). It is triggered by prodeath ligands like TNF α and requires the inhibition or loss of caspase-8. In vitro, necroptosis is typically induced by TNF α and a caspase inhibitor such as z-VAD.fmk or Q-VD-Oph. The latter are both broad-spectrum inhibitors that target caspase-8, caspase-9, and caspase-3/7 (Chauvier et al., 2007). If necroptotic stimuli induce mitochondrial damage and mtDNA release, our results would predict that pan-caspase inhibition would result in IFN production by dying cells. Previous studies suggest that necroptosis triggered by TNF α , z-VAD.fmk, and cycloheximide causes Bak/Bax-dependent mitochondrial damage (Irrinki et al., 2011) and that *Bak*^{-/-} *Bax*^{-/-} MEFs are partially protected from necroptotic death induced by TNF α , Q-VD-Oph, and a Smac mimetic (TSQ) (Moujalled et al., 2013). Given that, like TNF α , type I IFNs, in combination with caspase inhibition, can induce necroptosis (Dillon et al., 2014; Thapa et al., 2013), our data suggest the potential for feed-forward effects driven by mtDNA-induced IFN production during caspase-inhibited necroptosis.

Pharmacological Inhibition of Apoptotic Caspases

Several caspase inhibitors have undergone clinical trials. They include the caspase-1 inhibitors VX-740 and VX-765 (Belnacasan) and the pan-caspase inhibitors IDN-6556 (Emricasan) and GS-9450 (O'Brien and Dixit, 2009). To date, none have received approval, although IDN-6556 is in ongoing trials for a number of indications, including alcoholic hepatitis, hepatic impairment, and islet transplantation. Evidence that VX-765 can block the pyroptotic death of HIV-infected CD4 T cells (Doitsh et al., 2014) has added to the hope that caspase inhibitors may yet find clinical application. That study highlighted their potential as a new class of antiviral drugs that target the host, not the virus. Our findings support this notion from an entirely different angle, suggesting that apoptotic caspase inhibition might be an effective means of amplifying endogenous IFN production. It will be interesting to see whether patients treated with a pancaspase inhibitor exhibit elevations in IFN- β levels.

Caspases Negatively Regulate DAMP Signaling

The traditional classification of caspases as inflammatory or apoptotic has broken down in the last decade, as it has become

(J) Immunoprecipitation (IP) followed by PCR. Left, immunoblot of lysates taken from *Mcl1*^{-/-} MEFs transduced with an expression plasmid encoding FLAG-cGAS. Right, data represent the fold change (FC) in enrichment of DNA fragments using anti-FLAG or IgG (negative control) to coprecipitate DNA in untreated MEFs or MEFs treated with ABT-737 (1 μ M) and Q-VD-Oph (QVD) (30 μ M) for the indicated time. DNA fragments were amplified by real-time qPCR using eight primer pairs for mtDNA and two primer pairs for gDNA. The relative locations of the mtDNA amplicons are shown (data are combined from two, with three replicates). Means were compared between treated and untreated samples. See also Figures S5, S6, and S7.

Unless otherwise stated, means were compared using a two-tailed t test. Data represent the mean \pm SEM. *p \leq 0.05, **p \leq 0.01, and ***p \leq 0.005.

apparent that inflammatory caspases such as caspase-1 can kill and that apoptotic caspases can mediate non-death processes. Our findings demonstrate that caspase-9, -3, and -7 are essential negative regulators of mtDNA-induced DAMP signaling. Although apoptotic caspases have been previously reported to inactivate the DAMPs HMGB1 (Kazama et al., 2008) and IL-33 (Cayrol and Girard, 2009; Lüthi et al., 2009), the absence of autoinflammatory disease in caspase-9-deficient bone marrow chimeras suggests that, at least in the context of the hematopoietic system, apoptotic caspases are dispensable for their regulation. Alternatively, IFN- β may mask their effects. A very cogent review of this subject recently proposed that mammalian caspases “can be construed to act as either positive or negative regulators of inflammation” (Martin et al., 2012). Given that type I IFNs are pleiotropic in nature, possessing both inflammatory and anti-inflammatory properties (Prinz and Knobloch, 2012), our results indicate that the apoptotic caspases serve as both.

EXPERIMENTAL PROCEDURES

Experimental Animals

All mice were backcrossed for at least ten generations on a C57BL/6 background. *Apaf1*^{-/-} (Yoshida et al., 1998), *Bak*^{-/-} (Lindsten et al., 2000), *Bax*^{-/-} (Lindsten et al., 2000), *Casp9*^{-/-} (Kuida et al., 1998), *Casp3*^{-/-} (Kuida et al., 1996), *Casp7*^{-/-} (Lakhani et al., 2006), *Ifnar1*^{-/-} (Hwang et al., 1995), *Myd88*^{-/-} (Adachi et al., 1998), and *STING*^{-/-} (Jin et al., 2011) strains have been described. Animal procedures were approved by the Walter and Eliza Hall Institute Animal Ethics Committee.

Cell Death Assays

Cell death was induced by exposure to ABT-737 (AbbVie), Dexamethasone (Sigma), WEHI-539 (MedChemExpress), or Etoposide (Hospira). Where indicated, cells were preincubated for 1 hr with MRT-67307 (Sigma) and for 15–30 min with ABT-737, followed by continuous exposure to Q-VD-Oph (SM Biochemicals) or zVAD.fmk (R&D Systems). Cell viability was quantified by CellTiterGlo (Promega) or flow cytometric analysis of cells excluding 5 μ g/ml propidium iodide (PI) (Sigma) and, where indicated, cells also negative for AnnexinV-FITC (InvivoGen) binding using a FACSCalibur (BD) or LSRII (BD). Caspase activity was assayed by the addition of caspase3/7Glo (Promega) or by immunoblotting as described in the Extended Experimental Procedures.

Measurement of IFN- β

IFN- β protein was measured using the VeriKine-HS Human or Mouse Interferon Beta ELISA (PBL Assay Science).

Generation of mtDNA-Depleted ρ^0 Cells

MEFs were cultured in Dulbecco's modified Eagle's medium (DMEM) (GIBCO) supplemented with 4 mM L-glutamine, 4.5 g/l glucose, 10% FCS, 100 μ g/ml sodium pyruvate, and 50 μ g/ml uridine as described (Hashiguchi and Zhang-Akiyama, 2009). 100 ng/ml ethidium bromide was added to the medium for 6–10 days. qPCR evaluation of mtDNA content by qPCR, transfection with Poly(I:C)(HMW)/LyoVec (InvivoGen), and expression profiling are described in Extended Experimental Procedures.

Immunoprecipitation/PCR

Mcl1^{-/-} MEFs transduced with an expression plasmid (pMSCV-IRES-Hygro) encoding Flag-tagged mouse cGAS were treated with ABT-737 and Q-VD-Oph or left untreated. Following crosslinking of DNA and associated proteins, immunoprecipitation was performed with anti-FLAG M2 (Sigma) or mouse-IgG (BD) and coprecipitated DNA examined by real-time qPCR. Oligonucleotide sequences are listed in Extended Experimental Procedures.

Statistics

Data and statistical methods are expressed as outlined in figure legends. Standard statistical methods were performed using Prism software (GraphPad). Bonferroni post hoc test was used to correct for multiple testing.

ACCESSION NUMBERS

The Gene Expression Omnibus accession numbers for the microarray data sets reported in this paper are GSE57934 and GSE59972.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.036>.

AUTHOR CONTRIBUTIONS

M.J.W. and B.T.K. designed research, analyzed data and wrote the paper. M.J.W., K.M., R.M.L., M.E.R., and D.M. performed research and analyzed data. D.C.S.H., J.C.C., M.F.v.D., G.L., M.J.H., and S.B. contributed vital reagents and insightful discussions.

ACKNOWLEDGMENTS

The authors thank J. Corbin and L. Di Rago for excellent technical assistance; K. Franks, E. Lanera, K. McGregor, S. Ross, K. Stoev, L. Wilkins, and E. Kryan for outstanding animal husbandry; S. Monard for expert flow cytometry assistance; T. Mak for *Apaf1*^{+/-} mice; R. Flavell for *Casp3*^{+/-} *Casp7*^{+/-} mice; C. Thompson for *Bak*^{+/-} and *Bax*^{+/-} mice; P. Crack for *Ifnar1*^{+/-} mice; D. Fairlie and E. Lee for Bim variants; K. Aumann and C. De Nardo for manuscript assistance; and A. Strasser, J. Choi, S. Best, A. Rongvaux, and R. Flavell for insightful discussions. This work was supported by Project Grants (516725 and 575535), Program Grants (461219, 461221, and 1016647), Fellowships (D.C.S.H. and B.T.K.), and an Independent Research Institutes Infrastructure Support Scheme Grant (361646) from the Australian National Health and Medical Research Council; a scholarship from the Leukemia Foundation of Australia (M.J.W.), fellowships from the Cancer Council of Victoria (M.J.W. and D.M.); an Australian Postgraduate Award scholarship (K.M.); a fellowship from the Sylvia and Charles Viertel Foundation (B.T.K.); the Australian Cancer Research Fund; and a Victorian State Government Operational Infrastructure Support Grant.

Received: June 9, 2014

Revised: September 22, 2014

Accepted: November 10, 2014

Published: December 18, 2014

REFERENCES

- Adachi, O., Kawai, T., Takeda, K., Matsumoto, M., Tsutsui, H., Sakagami, M., Nakanishi, K., and Akira, S. (1998). Targeted disruption of the MyD88 gene results in loss of IL-1- and IL-18-mediated function. *Immunity* 9, 143–150.
- Barber, G.N. (2014). STING-dependent cytosolic DNA sensing pathways. *Trends Immunol.* 35, 88–93.
- Burguillos, M.A., Deierborg, T., Kavanagh, E., Persson, A., Hajji, N., Garcia-Quintanilla, A., Cano, J., Brundin, P., Englund, E., Venero, J.L., and Joseph, B. (2011). Caspase signalling controls microglia activation and neurotoxicity. *Nature* 472, 319–324.
- Cayrol, C., and Girard, J.-P. (2009). The IL-1-like cytokine IL-33 is inactivated after maturation by caspase-1. *Proc. Natl. Acad. Sci. USA* 106, 9021–9026.
- Chauvier, D., Ankri, S., Charriaud-Marlangue, C., Casimir, R., and Jacotot, E. (2007). Broad-spectrum caspase inhibitors: from myth to reality? *Cell Death Differ.* 14, 387–391.

- Civril, F., Deimling, T., de Oliveira Mann, C.C., Ablasser, A., Moldt, M., Witte, G., Hornung, V., and Hopfner, K.P. (2013). Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* 498, 332–337.
- Crawford, E.D., and Wells, J.A. (2011). Caspase substrates and cellular remodeling. *Annu. Rev. Biochem.* 80, 1055–1087.
- Crawford, E.D., Seaman, J.E., Agard, N., Hsu, G.W., Julien, O., Mahrus, S., Nguyen, H., Shimbo, K., Yoshihara, H.A., Zhuang, M., et al. (2013). The DegraBase: a database of proteolysis in healthy and apoptotic human cells. *Mol. Cell Proteomics* 12, 813–824.
- Dillon, C.P., Weinlich, R., Rodriguez, D.A., Cripps, J.G., Quarato, G., Gurung, P., Verbist, K.C., Brewer, T.L., Llambi, F., Gong, Y.-N., et al. (2014). RIPK1 blocks early postnatal lethality mediated by caspase-8 and RIPK3. *Cell* 157, 1189–1202.
- Doitsh, G., Galloway, N.L.K., Geng, X., Yang, Z., Monroe, K.M., Zepeda, O., Hunt, P.W., Hatano, H., Sowinski, S., Muñoz-Arias, I., and Greene, W.C. (2014). Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature* 505, 509–514.
- Domen, J., Cheshier, S.H., and Weissman, I.L. (2000). The role of apoptosis in the regulation of hematopoietic stem cells: Overexpression of Bcl-2 increases both their number and repopulation potential. *J. Exp. Med.* 191, 253–264.
- Ekert, P.G., Read, S.H., Silke, J., Marsden, V.S., Kaufmann, H., Hawkins, C.J., Gerl, R., Kumar, S., and Vaux, D.L. (2004). Apaf-1 and caspase-9 accelerate apoptosis, but do not determine whether factor-deprived or drug-treated cells die. *J. Cell Biol.* 165, 835–842.
- Enari, M., Sakahira, H., Yokoyama, H., Okawa, K., Iwamatsu, A., and Nagata, S. (1998). A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* 391, 43–50.
- Essers, M.A.G., Offner, S., Blanco-Bose, W.E., Waibler, Z., Kalinke, U., Duchosal, M.A., and Trumpp, A. (2009). IFN α activates dormant haematopoietic stem cells in vivo. *Nature* 458, 904–908.
- Hashiguchi, K., and Zhang-Akiyama, Q.M. (2009). Establishment of human cell lines lacking mitochondrial DNA. *Methods Mol. Biol.* 554, 383–391.
- Honarpour, N., Tabuchi, K., Stark, J.M., Hammer, R.E., Südhof, T.C., Parada, L.F., Wang, X., Richardson, J.A., and Herz, J. (2001). Embryonic neuronal death due to neurotrophin and neurotransmitter deprivation occurs independent of Apaf-1. *Neuroscience* 106, 263–274.
- Huang, Q., Li, F., Liu, X., Li, W., Shi, W., Liu, F.-F., O'Sullivan, B., He, Z., Peng, Y., Tan, A.C., et al. (2011). Caspase 3-mediated stimulation of tumor cell repopulation during cancer radiotherapy. *Nat. Med.* 17, 860–866.
- Hwang, S.Y., Hertzog, P.J., Holland, K.A., Sumarsono, S.H., Tymms, M.J., Hamilton, J.A., Whitty, G., Bertoncello, I., and Kola, I. (1995). A null mutation in the gene encoding a type I interferon receptor component eliminates anti-proliferative and antiviral responses to interferons alpha and beta and alters macrophage responses. *Proc. Natl. Acad. Sci. USA* 92, 11284–11288.
- Irrinki, K.M., Mallikankaraman, K., Thapa, R.J., Chandramoorthy, H.C., Smith, F.J., Jog, N.R., Gandhirajan, R.K., Kelsen, S.G., Houser, S.R., May, M.J., et al. (2011). Requirement of FADD, NEMO, and BAX/BAK for aberrant mitochondrial function in tumor necrosis factor alpha-induced necrosis. *Mol. Cell Biol.* 31, 3745–3758.
- Janzen, V., Fleming, H.E., Riedt, T., Karlsson, G., Riese, M.J., Lo Celso, C., Reynolds, G., Milne, C.D., Paige, C.J., Karlsson, S., et al. (2008). Hematopoietic stem cell responsiveness to exogenous signals is limited by caspase-3. *Cell Stem Cell* 2, 584–594.
- Jin, L., Hill, K.K., Filak, H., Mogan, J., Knowles, H., Zhang, B., Perraud, A.L., Cambier, J.C., and Lenz, L.L. (2011). MPYS is required for IFN response factor 3 activation and type I IFN production in the response of cultured phagocytes to bacterial second messengers cyclic-di-AMP and cyclic-di-GMP. *J. Immunol.* 187, 2595–2601.
- Kazama, H., Ricci, J.-E., Herndon, J.M., Hoppe, G., Green, D.R., and Ferguson, T.A. (2008). Induction of immunological tolerance by apoptotic cells requires caspase-dependent oxidation of high-mobility group box-1 protein. *Immunity* 29, 21–32.
- King, M.P., and Attardi, G. (1989). Human cells lacking mtDNA: repopulation with exogenous mitochondria by complementation. *Science* 246, 500–503.
- Kuida, K., Zheng, T.S., Na, S., Kuan, C., Yang, D., Karasuyama, H., Rakic, P., and Flavell, R.A. (1996). Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. *Nature* 384, 368–372.
- Kuida, K., Haydar, T.F., Kuan, C.Y., Gu, Y., Taya, C., Karasuyama, H., Su, M.S., Rakic, P., and Flavell, R.A. (1998). Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. *Cell* 94, 325–337.
- Labi, V., Bertele, D., Woess, C., Tischner, D., Bock, F.J., Schwemmers, S., Pahl, H.L., Geley, S., Kunze, M., Niemeyer, C.M., et al. (2013). Haematopoietic stem cell survival and transplantation efficacy is limited by the BH3-only proteins Bim and Bmf. *EMBO Mol. Med.* 5, 122–136.
- Lakhani, S.A., Masud, A., Kuida, K., Porter, G.A., Jr., Booth, C.J., Mehal, W.Z., Inayat, I., and Flavell, R.A. (2006). Caspases 3 and 7: key mediators of mitochondrial events of apoptosis. *Science* 311, 847–851.
- Lee, E.F., Czabotar, P.E., van Delft, M.F., Michalak, E.M., Boyle, M.J., Willis, S.N., Puthalakath, H., Bouillet, P., Colman, P.M., Huang, D.C., and Fairlie, W.D. (2008). A novel BH3 ligand that selectively targets Mcl-1 reveals that apoptosis can proceed without Mcl-1 degradation. *J. Cell Biol.* 180, 341–355.
- Li, P., Nijhawan, D., Budihardjo, I., Srinivasula, S.M., Ahmad, M., Alnemri, E.S., and Wang, X. (1997). Cytochrome c and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade. *Cell* 91, 479–489.
- Li, F., He, Z., Shen, J., Huang, Q., Li, W., Liu, X., He, Y., Wolf, F., and Li, C.Y. (2010a). Apoptotic caspases regulate induction of iPSCs from human fibroblasts. *Cell Stem Cell* 7, 508–520.
- Li, Z., Jo, J., Jia, J.M., Lo, S.C., Whitcomb, D.J., Jiao, S., Cho, K., and Sheng, M. (2010b). Caspase-3 activation via mitochondria is required for long-term depression and AMPA receptor internalization. *Cell* 141, 859–871.
- Lindsten, T., Ross, A.J., King, A., Zong, W.X., Rathmell, J.C., Shiels, H.A., Ulrich, E., Waymire, K.G., Mahar, P., Frauwirth, K., et al. (2000). The combined functions of proapoptotic Bcl-2 family members bak and bax are essential for normal development of multiple tissues. *Mol. Cell* 6, 1389–1399.
- Linkermann, A., and Green, D.R. (2014). Necroptosis. *N. Engl. J. Med.* 370, 455–465.
- Lüthi, A.U., Cullen, S.P., McNeela, E.A., Durie, P.J., Afonina, I.S., Sheridan, C., Brumatti, G., Taylor, R.C., Kersse, K., Vandenabeele, P., et al. (2009). Suppression of interleukin-33 bioactivity through proteolysis by apoptotic caspases. *Immunity* 31, 84–98.
- Marsden, V.S., O'Connor, L., O'Reilly, L.A., Silke, J., Metcalf, D., Ekert, P.G., Huang, D.C., Cecconi, F., Kuida, K., Tomaselli, K.J., et al. (2002). Apoptosis initiated by Bcl-2-regulated caspase activation independently of the cytochrome c/Apaf-1/caspase-9 apoptosome. *Nature* 419, 634–637.
- Martin, S.J., Henry, C.M., and Cullen, S.P. (2012). A perspective on mammalian caspases as positive and negative regulators of inflammation. *Mol. Cell* 46, 387–397.
- McIlwain, D.R., Berger, T., and Mak, T.W. (2013). Caspase functions in cell death and disease. *Cold Spring Harb. Perspect. Biol.* 5, a008656.
- Moujalled, D.M., Cook, W.D., Okamoto, T., Murphy, J., Lawlor, K.E., Vince, J.E., and Vaux, D.L. (2013). TNF can activate RIPK3 and cause programmed necrosis in the absence of RIPK1. *Cell Death Dis.* 4, e465.
- Nonomura, K., Yamaguchi, Y., Hamachi, M., Koike, M., Uchiyama, Y., Nakazato, K., Mochizuki, A., Sakaue-Sawano, A., Miyawaki, A., Yoshida, H., et al. (2013). Local apoptosis modulates early mammalian brain development through the elimination of morphogen-producing cells. *Dev. Cell* 27, 621–634.
- O'Brien, T., and Dixit, V.M. (2009). Drug discovery in apoptosis. In *Encyclopedia of Life Sciences*. Published online December 15, 2009. <http://dx.doi.org/10.1002/9780470015902.a0021587>.
- Oltersdorf, T., Elmore, S.W., Shoemaker, A.R., Armstrong, R.C., Augeri, D.J., Belli, B.A., Bruncko, M., Deckwerth, T.L., Dinges, J., Hajduk, P.J., et al. (2005). An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 435, 677–681.

- Opferman, J.T., Iwasaki, H., Ong, C.C., Suh, H., Mizuno, S., Akashi, K., and Korsmeyer, S.J. (2005). Obligate role of anti-apoptotic MCL-1 in the survival of hematopoietic stem cells. *Science* 307, 1101–1104.
- Oppenheim, R.W., Flavell, R.A., Vinsant, S., Prevette, D., Kuan, C.Y., and Rakic, P. (2001). Programmed cell death of developing mammalian neurons after genetic deletion of caspases. *J. Neurosci.* 21, 4752–4760.
- Orelia, C., and Dzierzak, E. (2007). Bcl-2 expression and apoptosis in the regulation of hematopoietic stem cells. *Leuk. Lymphoma* 48, 16–24.
- Paludan, S.R., and Bowie, A.G. (2013). Immune sensing of DNA. *Immunity* 38, 870–880.
- Prinz, M., and Knobloch, K.-P. (2012). Type I interferons as ambiguous modulators of chronic inflammation in the central nervous system. *Front. Immunol.* 3, 67.
- Sato, T., Onai, N., Yoshihara, H., Arai, F., Suda, T., and Ohteki, T. (2009). Interferon regulatory factor-2 protects quiescent hematopoietic stem cells from type I interferon-dependent exhaustion. *Nat. Med.* 15, 696–700.
- Shimada, K., Crother, T.R., Karlin, J., Dagvadorj, J., Chiba, N., Chen, S., Ram-anujan, V.K., Wolf, A.J., Vergnes, L., Ojcius, D.M., et al. (2012). Oxidized mitochondrial DNA activates the NLRP3 inflammasome during apoptosis. *Immunity* 36, 401–414.
- Sun, L., Wu, J., Du, F., Chen, X., and Chen, Z.J. (2013). Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* 339, 786–791.
- Suzuki, J., Denning, D.P., Imanishi, E., Horvitz, H.R., and Nagata, S. (2013). Xk-related protein 8 and CED-8 promote phosphatidylserine exposure in apoptotic cells. *Science* 341, 403–406.
- Thapa, R.J., Nogusa, S., Chen, P., Maki, J.L., Lerro, A., Andrade, M., Rall, G.F., Degterev, A., and Balachandran, S. (2013). Interferon-induced RIP1/RIP3-mediated necrosis requires PKR and is licensed by FADD and caspases. *Proc. Natl. Acad. Sci. U.S.A* 110, E3109–E3118.
- van Delft, M.F., Wei, A.H., Mason, K.D., Vandenberg, C.J., Chen, L., Czabotar, P.E., Willis, S.N., Scott, C.L., Day, C.L., Cory, S., et al. (2006). The BH3 mimetic ABT-737 targets selective Bcl-2 proteins and efficiently induces apoptosis via Bak/Bax if Mcl-1 is neutralized. *Cancer Cell* 10, 389–399.
- van Delft, M.F., Smith, D.P., Lahoud, M.H., Huang, D.C.S., and Adams, J.M. (2010). Apoptosis and non-inflammatory phagocytosis can be induced by mitochondrial damage without caspases. *Cell Death Differ.* 17, 821–832.
- Wei, M.C., Zong, W.X., Cheng, E.H., Lindsten, T., Panoutsakopoulou, V., Ross, A.J., Roth, K.A., MacGregor, G.R., Thompson, C.B., and Korsmeyer, S.J. (2001). Proapoptotic BAX and BAK: a requisite gateway to mitochondrial dysfunction and death. *Science* 292, 727–730.
- Yi, C.H., and Yuan, J. (2009). The Jekyll and Hyde functions of caspases. *Dev. Cell* 16, 21–34.
- Yoshida, H., Kong, Y.Y., Yoshida, R., Elia, A.J., Hakem, A., Hakem, R., Penninger, J.M., and Mak, T.W. (1998). Apaf1 is required for mitochondrial pathways of apoptosis and brain development. *Cell* 94, 739–750.
- Youle, R.J., and Strasser, A. (2008). The BCL-2 protein family: opposing activities that mediate cell death. *Nat. Rev. Mol. Cell Biol.* 9, 47–59.

Apoptotic Caspases Prevent the Induction of Type I Interferons by Mitochondrial DNA

Anthony Rongvaux,¹ Ruaidhrí Jackson,¹ Christian C.D. Harman,¹ Tuo Li,² A. Phillip West,³ Marcel R. de Zoete,^{1,7} Youtong Wu,² Brian Yordy,¹ Saquib A. Lakhani,^{1,8} Chia-Yi Kuan,⁵ Tadatsugu Taniguchi,⁶ Gerald S. Shadel,^{3,4} Zhijian J. Chen,^{2,7} Akiko Iwasaki,^{1,7} and Richard A. Flavell^{1,7,*}

¹Department of Immunobiology, Yale University School of Medicine, New Haven, CT 06520, USA

²Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

³Department of Pathology

⁴Departments of Genetics

Yale University School of Medicine, New Haven, CT 06520, USA

⁵Department of Pediatrics, Division of Neurology and The Center for Neurodegenerative Diseases (CND), Emory University School of Medicine, Atlanta, GA 30322, USA

⁶Department of Molecular Immunology, Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan

⁷Howard Hughes Medical Institute

⁸Present address: Department of Pediatrics, University of South Dakota Sanford School of Medicine, Sioux Falls, SD 57117, USA

*Correspondence: richard.flavell@yale.edu
<http://dx.doi.org/10.1016/j.cell.2014.11.037>

SUMMARY

The mechanism by which cells undergo death determines whether dying cells trigger inflammatory responses or remain immunologically silent. Mitochondria play a central role in the induction of cell death, as well as in immune signaling pathways. Here, we identify a mechanism by which mitochondria and downstream proapoptotic caspases regulate the activation of antiviral immunity. In the absence of active caspases, mitochondrial outer membrane permeabilization by Bax and Bak results in the expression of type I interferons (IFNs). This induction is mediated by mitochondrial DNA-dependent activation of the cGAS/STING pathway and results in the establishment of a potent state of viral resistance. Our results show that mitochondria have the capacity to simultaneously expose a cell-intrinsic inducer of the IFN response and to inactivate this response in a caspase-dependent manner. This mechanism provides a dual control, which determines whether mitochondria initiate an immunologically silent or a proinflammatory type of cell death.

INTRODUCTION

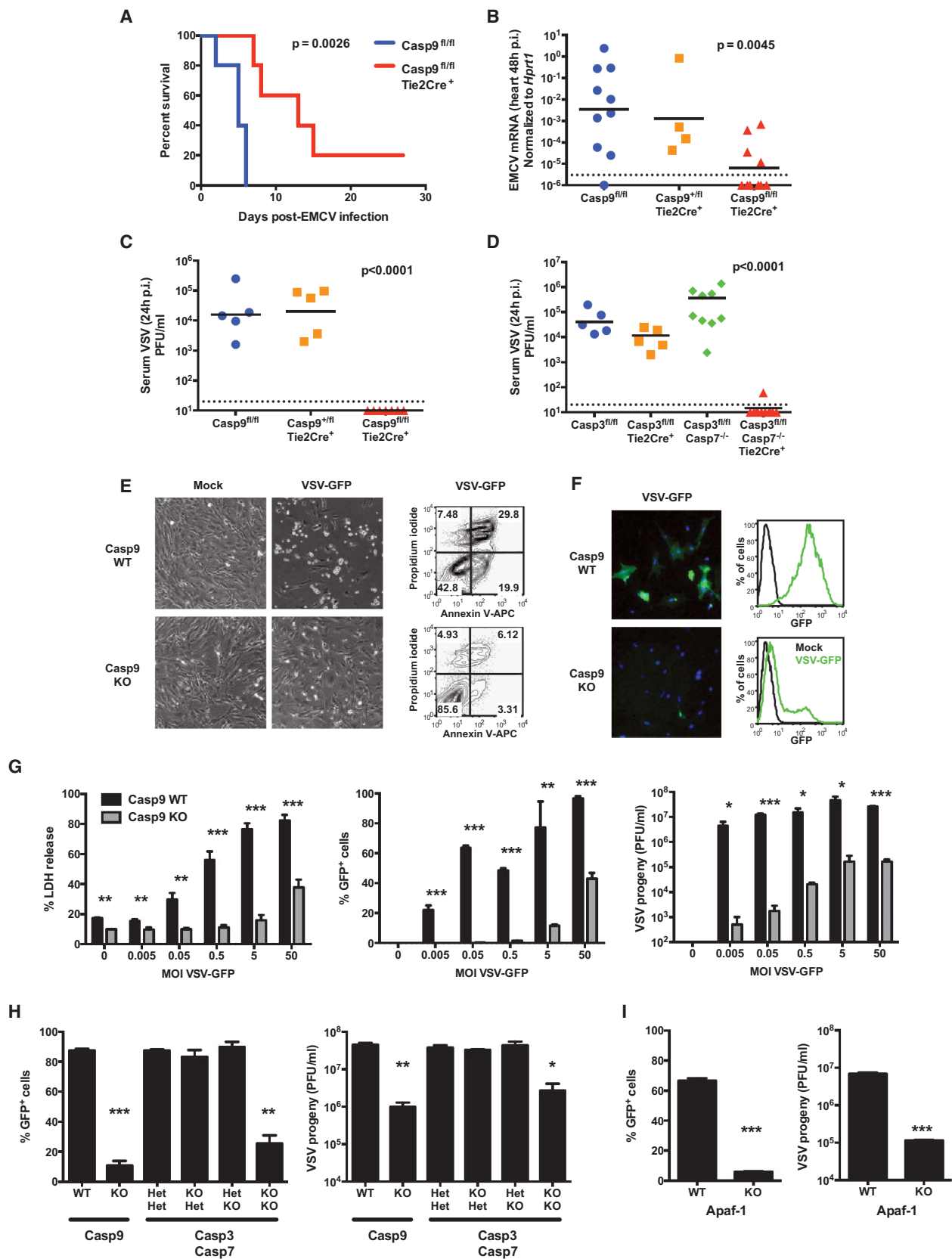
Multicellular organisms are constantly exposed to the threat of viral infections. As a response, vertebrates have evolved several mechanisms of antiviral defense. These mechanisms include the production of type I interferons (IFNs) (Stetson and Medzhitov, 2006) and the suicide of infected cells (Upton and Chan, 2014).

Type I IFNs (IFN α and IFN β) are cytokines of major importance for the innate antiviral response (Stetson and Medzhitov, 2006).

They are produced after recognition of viral nucleic acids by toll-like receptors (TLRs) or by cytoplasmic proteins such as RIG-I-like receptors (RLRs) or the cyclic GMP-AMP synthase (cGAS) (Cai et al., 2014; Kawai and Akira, 2011; Loo and Gale, 2011). After their secretion, type I IFNs bind to the type I IFN receptor (IFNAR) in an autocrine and paracrine manner. This signal induces the expression of hundreds of interferon-stimulated genes (ISGs) in the responding cell (Schneider et al., 2014). Overall, ISGs have the capacity to interfere with every step of viral replication, and, as a consequence, the IFN response results in the establishment of a cellular state of viral resistance.

The programmed death of infected cells limits the possibility for viruses to subvert the cellular machinery for their own replication (Best, 2008; Yatim and Albert, 2011). One of the best-described mechanisms of programmed cell death is apoptosis, which is mediated through the activation of members of the caspase family of proteases (Fuchs and Steller, 2011; Kumar, 2007; Taylor et al., 2008). The mitochondrial pathway of apoptosis is induced in response to cellular stress. It is regulated by the activities of pro- and antiapoptotic members of the Bcl-2 family, which control the formation of the Bax/Bak channel that results in mitochondrial outer membrane permeabilization (MOMP) (Chipuk et al., 2010; Tait and Green, 2010; Youle and Strasser, 2008). Following MOMP, mitochondrial proteins, including cytochrome c, are released in the cytosol. Together with Apaf-1 and caspase-9, cytosolic cytochrome c forms a protein complex called the apoptosome, which induces the activation of caspase-9 (Jiang and Wang, 2004; Riedl and Salvesen, 2007). The downstream effector caspases-3 and -7 are cleaved and activated by caspase-9, triggering a cascade of proteolytic events that culminates in the demise of the cell through apoptosis (Kroemer et al., 2009).

Although caspases are key mediators of apoptotic cell death (Kumar, 2007), multiple mechanisms of caspase-independent cell death exist (Chipuk and Green, 2005; Tait et al., 2014; Vanden Berghe et al., 2014). The discovery of a broad diversity of



(legend on next page)

nonapoptotic death pathways has led to a re-evaluation of caspases as essential mediators of cell death. An appealing hypothesis to reconcile the evolutionary conservation of proapoptotic caspase signaling with the existence of multiple, and potentially redundant, death-inducing pathways is that caspase-dependent apoptosis is unique in its capacity to induce an immunologically silent form of cell death, whereas other types of cell death have proinflammatory or immunostimulatory properties (Martin et al., 2012; Tait et al., 2014). Indeed, necrotic cell death results in the release of molecules with proinflammatory properties, collectively termed damage-associated molecular patterns (DAMPs) or alarmins (Kroemer et al., 2013). Mounting evidence demonstrates that several DAMPs can be inactivated in a caspase-dependent manner during apoptosis, supporting the importance of caspases in maintaining cell death as immunologically silent. However, it is probable that a large spectrum of caspase-dependent mechanisms of immune regulation remain to be discovered (Martin et al., 2012).

In this study, we identify an unsuspected mechanism by which the mitochondrial events of apoptosis actively trigger the initiation of a cell-intrinsic immune response, mediated by the expression of type I IFNs. Proapoptotic caspases, activated simultaneously by mitochondria, are required to inhibit that response and to maintain apoptosis immunologically silent. Therefore, mitochondria and caspases play a crucial role not only in the decision of the cell to live or to commit suicide but also on the decision to die in an inflammatory or immunologically silent manner.

RESULTS

Intrinsic Apoptosis Deficiency Confers Resistance to Viral Infection

As mice with genetic deficiencies in the intrinsic pathway of apoptosis die perinatally (Hakem et al., 1998; Kuida et al., 1998; Lakhani et al., 2006; Yoshida et al., 1998), we generated mice with a floxed caspase-9 allele or a floxed caspase-3 allele (Figure S1A available online). With the initial objective of studying the role of the intrinsic pathway of apoptosis in immune cells, we crossed *Casp9^{fl/fl}* mice and *Casp3^{fl/fl} Casp7^{-/-}* mice with Tie2-Cre(E+H) (Koni et al., 2001; Lakhani et al., 2006) to obtain mice with endothelial/hematopoietic tissue-specific deletion of the

respective floxed alleles (Figures S1B–S1F). We observed that *Casp9^{fl/fl} Tie2-Cre⁺* and *Casp3^{fl/fl} Casp7^{-/-} Tie2-Cre⁺* mice were highly resistant to viral infection in comparison to littermate controls. Indeed, lethality following intraperitoneal infection with encephalomyocarditis virus (EMCV, 2×10^3 TCID₅₀), which was observed in control mice 6 days after infection, was delayed in *Casp9^{fl/fl} Tie2-Cre⁺* mice (Figure 1A). The prolonged survival and resistance to EMCV infection was associated with lower viral loads in the heart 2 days after infection, with undetectable expression of the EMCV genome in half of the *Casp9^{fl/fl} Tie2-Cre⁺* mice (Figure 1B). Furthermore, the deletion of caspase-9 or of both caspases-3 and -7 resulted in undetectable viral titers of vesicular stomatitis virus (VSV, 10^6 PFU) after intranasal infection (Figures 1C and 1D), highlighting a potent antiviral state in vivo in proapoptotic caspase-deficient animals.

To determine whether this phenotype could be recapitulated in vitro, primary mouse embryonic fibroblasts (MEFs) isolated from caspase-9 knockout (KO) (Kuida et al., 1998) from caspase-3/-7 double-KO (Lakhani et al., 2006) and from Apaf-1 KO mice (Yoshida et al., 1998) were infected with VSV. We observed that caspase-9-deficient cells (*Casp9* KO) were only modestly affected by the infection with a recombinant strain of VSV expressing the green fluorescent protein (VSV-GFP), whereas wild-type (WT) cells derived from littermate embryos (*Casp9* WT) showed the typical phenotype of infected cells (cell rounding, detachment, and death; Figure 1E). Fluorescence microscopy and flow cytometry analysis showed that only a small fraction of *Casp9* KO cells expressed virus-encoded GFP (Figure 1F). To further substantiate this observation, *Casp9* WT and KO MEFs were infected with VSV-GFP at various multiplicities of infection (MOI). By measuring cell death (percentage of LDH released) and viral infection and replication (GFP expression and plaque-forming units), we observed that caspase-9 deficiency significantly reduced the susceptibility of cells to VSV infection at all MOIs tested (Figure 1G). *Casp9* KO MEFs also displayed increased resistance to infection by EMCV (Figure S2A) or by herpes simplex virus type 2 (HSV-2) (Figure S2B). Similarly, *Casp3/7* double-KO and Apaf-1 KO MEFs showed resistance to VSV infection comparable to that observed in *Casp9* KO (Figures 1H and 1I). These results demonstrate that deficiency in the intrinsic pathway of apoptosis, downstream of

Figure 1. Loss of the Intrinsic Pathway of Apoptosis Enhances Resistance to Viral Infection

(A and B) *Casp9^{fl/fl} Tie2-Cre⁺* and control mice were infected intraperitoneally with EMCV (2×10^3 TCID₅₀), and the survival was monitored ($n = 5$ mice/group, p value calculated by Mantel-Cox test) (A); or the mice were sacrificed 48 hr postinfection (p.i.), and viral loads in the heart were measured by real-time RT-PCR ($n = 4$ –10 mice/group, combined from three independent experiments, p value calculated by one-way ANOVA) (B). Each symbol represents an individual mouse, and the black horizontal bars indicate geometric means. The dashed line indicates the limit of detection of the assay.

(C and D) *Casp9^{fl/fl} Tie2-Cre⁺* (C), *Casp3^{fl/fl} Casp7^{-/-} Tie2-Cre⁺* (D), and control mice were infected intranasally with VSV (10^6 PFU) and sacrificed 24 hr later. Viral loads were measured in the plasma by plaque-forming assay ($n = 5$ –7 mice/group, combined from at least two independent experiments, p value calculated by one-way ANOVA).

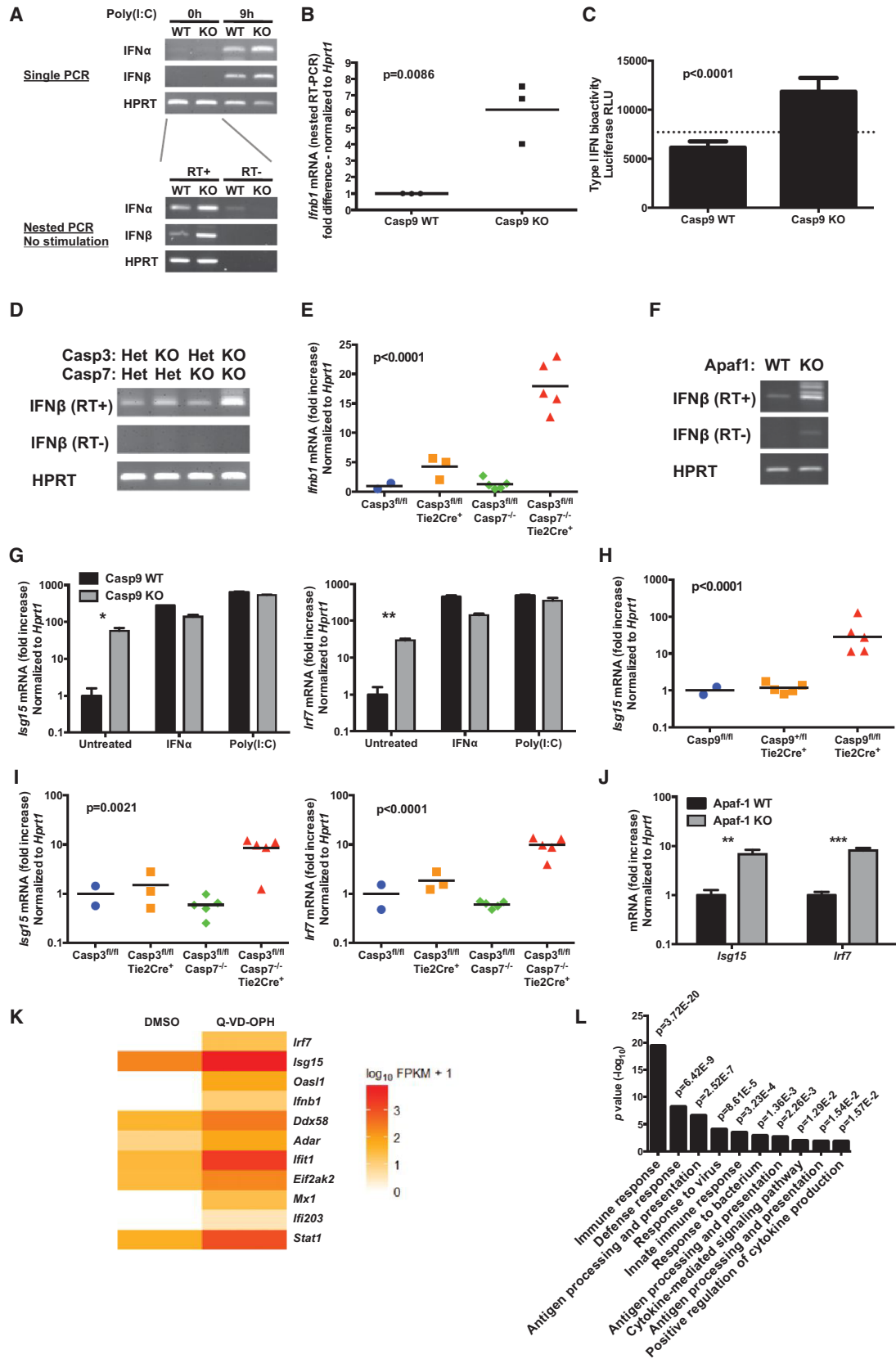
(E and F) *Casp9* WT and KO primary MEFs were infected in vitro with VSV-GFP (MOI = 0.5) and analyzed 24 hr later (E). The expression of virus-encoded GFP was analyzed by fluorescence microscopy (green, GFP; blue, counter-staining of nuclei with DAPI) or by flow cytometry (F).

(G) *Casp9* WT and KO primary MEFs were infected with the indicated MOI of VSV-GFP and were assessed 24 hr later for cell death with LDH release assay (left), expression of GFP (middle), and viral progeny production by plaque assay (right). Results are presented as mean \pm SD of triplicates, representative of at least three independent experiments.

(H and I) *Casp3/7* double-KO (H) or Apaf-1 KO (I) and respective control primary MEFs were infected with VSV-GFP (MOI = 0.5), and GFP expression and viral progeny were measured as in (G) (mean \pm SD of duplicates, representative of two experiments).

* $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ (two-tailed unpaired Student's t test, compared to respective WT or Het/Het control).

See also Figures S1 and S2.



(legend on next page)

mitochondria, confers a strong broad-spectrum resistance to infection by RNA and DNA viruses, both *in vivo* and *in vitro*.

Constitutive Activation of the Type I IFN Response

Type I IFNs are critical determinants of the cellular susceptibility to viral infections. They are constitutively expressed at low levels, and these steady-state IFNs have profound physiological effects on homeostasis through tonic signaling in the absence of acute infection (Gough et al., 2012; Taniguchi and Takaoka, 2001). We measured the baseline expression levels of type I IFNs using a highly sensitive nested RT-PCR. We observed a modest increase in the basal levels of the mRNA encoding both IFN α and IFN β in Casp9 KO MEFs compared to WT controls (Figure 2A) in the absence of any stimulation. We confirmed this result using a quantitative nested real-time PCR (Figure 2B), as well as a type I IFN bioactivity assay (Figure 2C). Increased steady-state type I IFN mRNA was also induced *in vitro* and *in vivo* by Casp3/7 double deficiency (Figures 2D and 2E), as well as by the absence of Apaf-1 in MEFs (Figure 2F).

In addition, we observed that interferon-stimulated genes (ISGs) were constitutively expressed at elevated levels *in vitro* and *in vivo* in Casp9 KO (Figures 2G, 2H, S3A, and S3B), in Casp3/7 double-KO (Figure 2I) and in Apaf-1 KO cells (Figure 2J) in the absence of any stimulation. The maximal expression level of these ISGs, as induced by IFN α or intracellular poly(I:C), was comparable between Casp9 WT and KO cells.

To determine whether the pharmacological inhibition of caspases could recapitulate the phenotype caused by genetic deficiencies, we treated WT MEFs with broad-spectrum inhibitors of caspases (Z-VAD-fmk, Boc-D-fmk, and Q-VD-OPH). These inhibitors induced an increased expression of ISGs (Figures 2K, S3C, and S3D), similar to the effect of caspase or Apaf-1 deficiency. Surprisingly, a gene ontology analysis of the genes differentially expressed between WT cells treated with dimethyl sulfoxide (DMSO) or Q-VD-OPH revealed a highly significant overrepresentation of pathways related to immune re-

sponses (Figure 2L). Although this transcriptional analysis does not take into account the direct proteolytic effects of caspases, it nevertheless reveals a profound effect of caspase inhibition on immune function.

Next, we compared the transcriptional changes induced by caspase inhibition in WT and IFNAR1 KO cells, which lack a critical subunit of the receptor for IFN α/β (Müller et al., 1994). We observed that the absence of the IFNAR receptor abrogated the transcriptional response of the cells to caspase inhibition by Q-VD-OPH (Figures 3A and S3D and Table S1), demonstrating the role of type I IFNs in this response.

To demonstrate that ISG expression and viral resistance in Casp9 KO cells was also due to type I IFNs, supernatants from confluent unstimulated Casp9 WT and KO MEFs were transferred to cultures of WT MEFs for 24 hr. The conditioned supernatant from Casp9 KO cells induced an increase in the expression of ISGs by WT MEFs (Figure 3B) to levels similar to those measured in Casp9 KO cells (compare with Figure 2G). Next, we pretreated WT cells with conditioned supernatants collected from Casp9 WT and KO cells in the absence or presence of neutralizing anti-IFN α/β antibodies. The cells were then washed and infected with VSV-GFP. The supernatants from Casp9 KO cell cultures conferred resistance to VSV infection in WT cells, and this effect was completely abolished by the presence of anti-IFN α/β neutralizing antibodies in the conditioned media during the pretreatment of the cells (Figure 3C). Similarly, conditioned supernatants from Casp9 KO cells failed to confer resistance to viral infection when used to pretreat IFNAR1 KO MEFs (Figure S4). These results demonstrate that the ISG-inducing activity and the resistance to VSV infection are mediated by the elevated concentrations of type I IFNs in the supernatant of Casp9 KO cells.

To further confirm this result, we generated Casp9/IFNAR1 double-KO MEFs. Like Casp9 KO cells, Casp9/IFNAR1 double-KO cells expressed increased steady-state levels of IFN α/β (Figure 3D), and their supernatant contained ISG-inducing

Figure 2. Inhibition of Intrinsic Apoptosis Activates the IFN Response

- (A) The expression of IFN α and IFN β mRNA was determined by RT-PCR in Casp9 WT and KO primary MEFs. Top: single RT-PCR on untreated cells and on cells transfected with poly(I:C) as a positive control. Bottom: nested RT-PCR on untreated cells (RT+, RNA reverse transcribed in cDNA; RT–, no reverse transcription).
- (B) The steady-state expression of IFN β mRNA expression in unstimulated primary MEFs was quantified by nested real-time RT-PCR. Each dot represents an independent experiment; p value: two-tailed unpaired Student's t test.
- (C) Type I IFN bioactivity in the culture supernatant of unstimulated MEFs was measured using an ISRE-Luc reporter cell line (mean \pm SD of six replicates, representative of two independent experiments; p value: two-tailed unpaired Student's t test; the dashed line indicates background from untreated reporter cells).
- (D) Nested RT-PCR amplification of steady-state IFN β in Casp3/7 double-deficient and control primary MEFs.
- (E) IFN β mRNA expression measured by real-time RT-PCR in Casp3/7-deficient and control spleen cells (n = 2–5 mice/genotype; p value calculated by one-way ANOVA).
- (F) Nested RT-PCR amplification of steady-state IFN β in Apaf-1 WT and KO primary MEFs.
- (G) The expression of selected ISGs in Casp9 WT and KO primary MEFs was measured by real-time RT-PCR. IFN α and intracellular poly(I:C) were used as positive controls (mean \pm SD of duplicates, representative of at least five independent experiments). *p < 0.05, **p < 0.01, and ***p < 0.001; two-tailed unpaired Student's t test.
- (H and I) ISG mRNA expression was measured by real-time RT-PCR in Casp9-deficient and control white blood cells (H) or in Casp3/7 double-deficient and control spleen cells (I) (n = 2–5 mice/genotype; p value: one-way ANOVA).
- (J) ISG mRNA expression measured by real-time RT-PCR in Apaf-1 WT and KO primary MEFs (mean \pm SD of triplicates, representative of three independent experiments; p value calculated as in G).
- (K) Heatmap of the expression of IFN β and selected ISGs in WT primary MEFs stimulated for 48 hr with vehicle (DMSO) or with the caspase inhibitor Q-VD-OPH (10 μ M).
- (L) Gene Ontology analysis of the pathways overrepresented among genes differentially expressed between WT primary MEFs stimulated with vehicle or with Q-VD-OPH.

See also Figures S3 and S5 and Table S1.

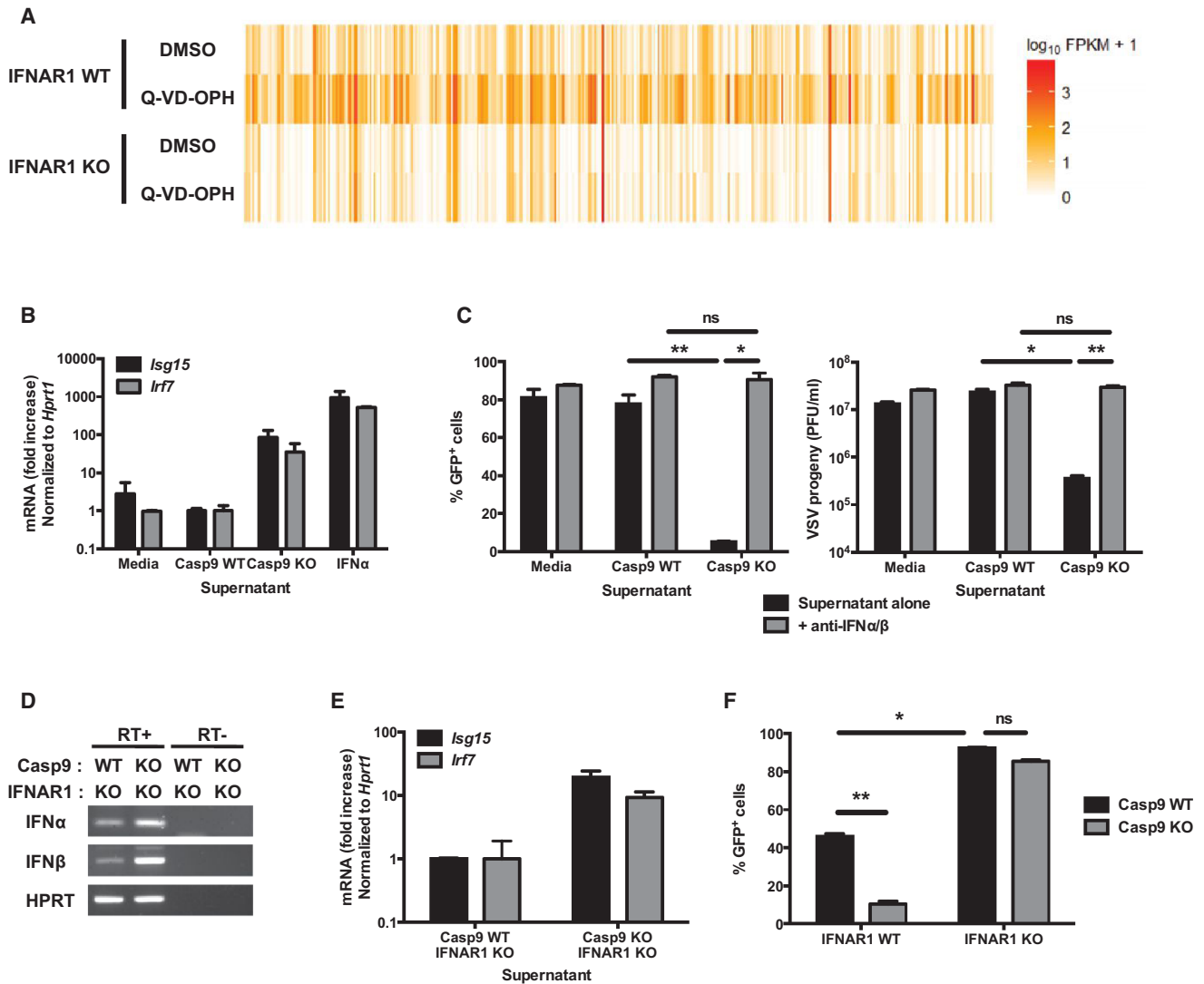


Figure 3. Expression of ISGs and Antiviral Resistance Is Mediated by Type I IFNs

(A) Heatmap of the expression of genes differentially expressed ($q < 0.05$, fold difference ≥ 5) in IFNAR1 WT and KO primary MEFs stimulated for 48 hr with vehicle (DMSO) or with the caspase inhibitor Q-VD-OPH (10 μ M) and determined by RNA sequencing on duplicate samples.

(B) Culture supernatants of confluent cultures of Casp9 WT and KO primary MEFs were collected. WT primary MEFs were then incubated for 16 hr in the presence of these supernatants or of recombinant IFN α (50 U/ml), and the expression level of selected ISGs was measured by real-time RT-PCR (mean \pm SD of duplicates; representative of two independent experiments).

(C) WT primary MEFs were incubated for 16 hr with conditioned supernatants from Casp9 WT or KO MEFs in the presence or absence of anti-IFN α and anti-IFN β neutralizing antibodies (300 NU/ml each). The cells were then washed, infected with VSV-GFP (MOI = 0.5, 24 hr), and the expression of GFP (left) and viral progeny production (right) was measured (mean \pm SD of triplicates, representative of three independent experiments).

(D) The expression of IFN α and IFN β mRNA was detected by nested RT-PCR in unstimulated Casp9 WT/IFNAR1 KO and Casp9 KO/IFNAR1 KO primary MEFs (RT+, RNA reverse transcribed in cDNA; RT-, no reverse transcription).

(E) WT primary MEFs were incubated for 16 hr with conditioned media from Casp9/IFNAR1 WT/KO or KO/KO MEFs, and the expression levels of ISGs were measured by real-time RT-PCR.

(F) Casp9/IFNAR1 double-KO and control primary MEFs were infected with VSV-GFP (MOI = 0.5, 24 hr), and the expression of GFP was measured by flow cytometry (mean \pm SD of duplicates, representative of three experiments).

* $p < 0.05$ and ** $p < 0.01$; ns, not significant; pairwise comparisons following two-way ANOVA (C and F).

See also Figure S4 and Tables S1 and S2.

activity (Figure 3E). However, in the absence of IFNAR1, Casp9 WT and KO cells were equally susceptible to VSV infection (Figure 3F). As Casp9/IFNAR1 double-KO cells are deficient in

apoptosis but are nevertheless susceptible to VSV, this result shows that viral resistance is not a direct consequence of defective cell death.

Caspase-9-deficient mice die during embryonic development or shortly after birth, and this phenotype has been attributed to apoptosis defects in the developing brain (Hakem et al., 1998; Kuida et al., 1998). However, we wanted to determine whether the constitutive activation of the IFN response and high expression of ISGs could contribute to this lethality. To this end, we compared the viability of caspase-9 KO mice in pre- and post-natal life in the presence or absence of IFNAR1 (Table S2). The absence of IFNAR1 did not rescue the embryonic lethality, showing that constitutive type I IFNs/ISGs expression is not responsible.

Aberrant expression of type I IFNs is the cause of several autoimmune disorders (Stetson, 2009). Surprisingly, however, despite constitutive expression of type I IFNs and ISGs, conditional Casp9 KO or Casp3/7 double-KO mice did not show any increase in total serum immunoglobulin or in antinuclear antibodies, two diagnostic characteristics of autoimmune diseases (Figure S5). We speculate that this absence of autoimmunity despite constitutive IFN response is due to pleiotropic functions of caspases and probable functional deficiencies in other mechanisms involved in the development of (auto)immunity.

Taken together, these observations demonstrate that, in the absence of a functional pathway of intrinsic apoptosis, an increased expression of steady-state type I IFNs is sufficient to induce ISG expression, and viral resistance is established. Such unexpected findings raise the intriguing questions as to what ligands and mechanisms govern type I IFN response in dying cells, how healthy cells contain unwanted IFN production, and finally, by what means proapoptotic caspases affect these processes.

Bax/Bak-Dependent Induction of Type I IFNs

The intrinsic pathway of apoptosis is activated upon MOMP by the Bax/Bak channel (Jiang and Wang, 2004). We thus wanted to determine whether mitochondria and Bax/Bak-dependent permeabilization were also involved in regulating the IFN response. Unlike deficiency in Apaf-1 or caspases, the absence of Bax and Bak (Figures 4A and 4B) or the overexpression of the antagonist protein Bcl-2 (Figure S6A), although preventing cell death induced by viral infection (Figures 4A, 4B, S6B, and S6C), did not confer any resistance to viral infection (Figures 4B and S6C). This observation could suggest that Apaf-1 and caspases regulate IFN expression independently of mitochondrial events. However, further investigations revealed a more subtle role of Bax and Bak in this process. Indeed, unlike WT cells, Bax/Bak-deficient cells treated with the caspase inhibitor Q-VD-OPH failed to induce the expression of ISGs (Figure 4C). This observation suggests that Bax/Bak are actually required for the induction of the IFN response in the absence of active caspases. As Bax/Bak-dependent permeabilization of mitochondrial outer membrane occurs in cultures of unstimulated cells in only a small percentage of dying cells or in cells undergoing incomplete MOMP, we hypothesized that the pharmacological inhibition of Bcl-2 could favor Bax/Bak-dependent MOMP and amplify the IFN response caused by caspase deficiency or inhibition. Consistently, we observed an induction of the expression of IFN β in Casp9 KO cells treated with the Bcl-2 inhibitor ABT-737 (Oltersdorf et al., 2005) (Figure 4D). In contrast, Casp9 WT cells did not express IFN β in response to Bcl-2 inhibition, showing the

role of caspases in regulating this response. As the constitutive expression of ISGs in Casp9 KO (Figures 2 and S3) could contribute to the response, we next treated WT cells with a combination of the Bcl-2 inhibitor ABT-737 and the caspase inhibitor Q-VD-OPH. The combination of Bcl-2/caspase inhibition (ABT-737 + Q-VD-OPH) resulted in a robust expression of IFN β after 3–6 hr of stimulation (Figure 4E). Other cytokines, such as IL-6 and TNF α , were only moderately induced (Figure S6D). The induction of IFN β mRNA by Bcl-2/caspase coinhibition was entirely dependent on the presence of Bax/Bak (Figure 4F). The treatment of human PBMCs with ABT-737 and Q-VD-OPH also induced high expression of IFN β (Figure 4G), showing that the mechanism of Bax/Bak-dependent caspase-regulated induction of type I IFN is conserved between species. Importantly, inhibiting caspases in the context of cell-extrinsic, caspase-8-dependent apoptosis induction did not induce the expression of IFN β , showing the specificity of this process for mitochondria-dependent apoptosis (Figure S6E).

Taken together, these results uncover an immunomodulatory role for the mitochondria in innate immunity, a process tightly regulated by proapoptotic caspases in which Bax/Bak-induced MOMP facilitates the release of a mitochondrial factor with the capacity to stimulate type I IFN expression and promote viral resistance.

Activation of the cGAS/STING Pathway

To identify the putative mitochondrial factor that induces type I IFN expression in response to Bcl-2/caspases coinhibition, we first determined which interferon-inducing pathway is involved. We used two criteria to determine the involvement of a candidate sensor or signaling molecule in this process: (1) the activation of this candidate factor after ABT-737 + Q-VD-OPH treatment and (2) the absence of IFN β expression, in response to the same treatment, in cells lacking the candidate factor.

Interferon regulatory factors (IRFs), and in particular IRF-3 and IRF-7, are major transcription factors required for the expression of type I IFNs (Tamura et al., 2008), and they are activated by the upstream kinase TBK1. Consistent with the expression of IFN β , the inhibition of Bcl-2 and caspases (ABT-737 + Q-VD-OPH) induced the phosphorylation of TBK1 and IRF-3 in Bax/Bak-sufficient cells, but not in Bax/Bak KO cells (Figure 5A). The phosphorylation of TBK1 and IRF-3 after transfection of HT-DNA (hering testes DNA, a stimulator of the interferon response that serves as a positive control) was not affected by the absence of Bax/Bak. The expression of IFN β in response to ABT-737 + Q-VD-OPH was completely abrogated in IRF-3/7 double-deficient cells (Figure 5B). These results demonstrate the critical involvement of TBK1 and IRFs in Bax/Bak-dependent caspase-regulated type I IFN induction. Two intracellular pathways converge on the TBK1/IRFs-dependent transcription of type I IFNs: the RLR/MAVS-dependent pathway activated by intracellular viral RNA (Loo and Gale, 2011) and the cGAS/STING pathway of cytosolic DNA recognition (Cai et al., 2014). MAVS deficiency did not affect the response to ABT-737 + Q-VD-OPH (Figure 5C), excluding a role of the cytosolic RNA recognition pathway. In contrast, the Bax/Bak-dependent, caspase-regulated IFN production was entirely dependent on the cGAS/STING pathway of cytosolic DNA recognition. The cGAS/STING

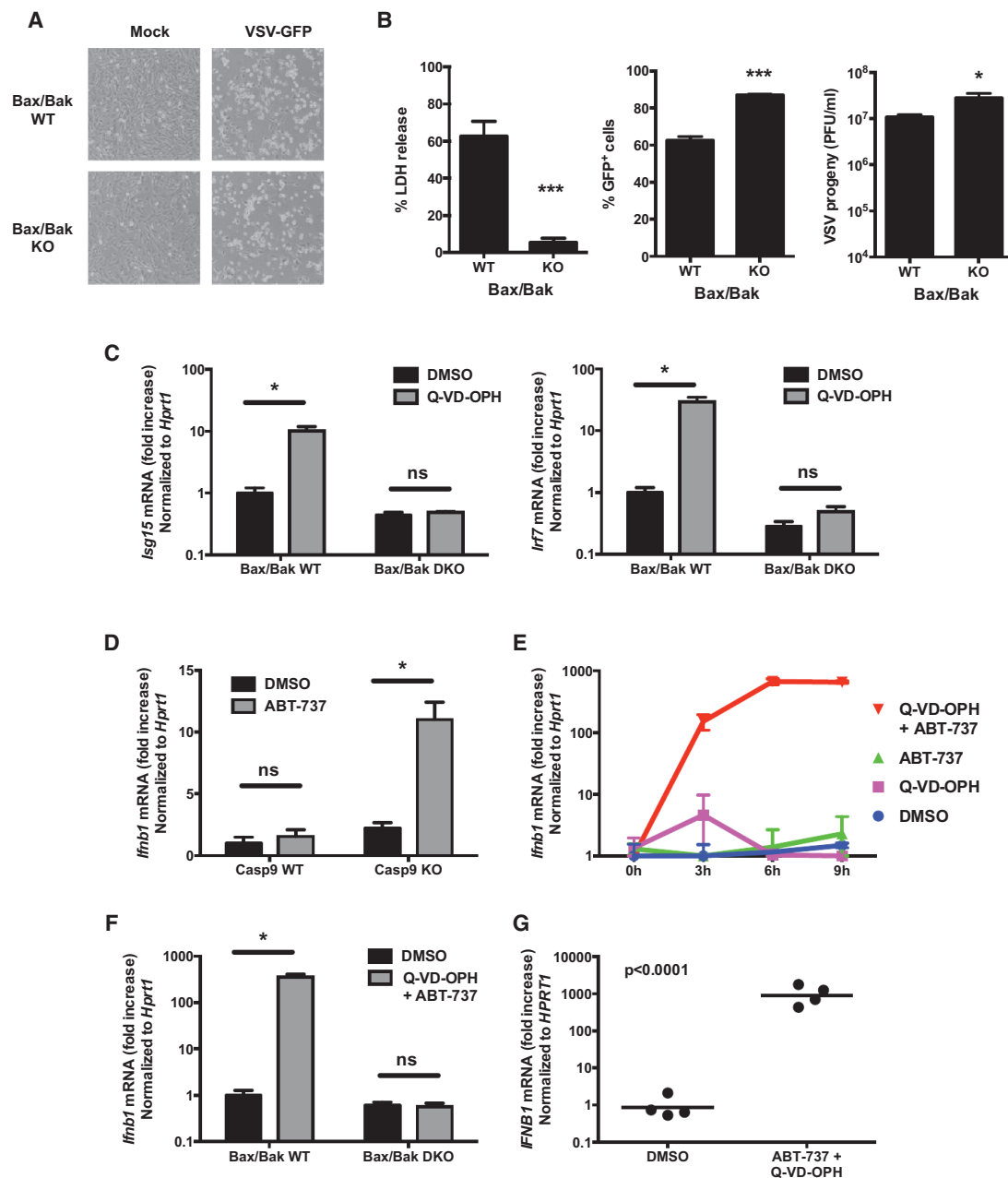


Figure 4. Bax/Bak-Dependent Induction of the IFN Response in the Absence of Active Caspases

(A and B) Bax/Bak double-KO and control immortalized MEFs were infected with VSV-GFP (MOI = 0.5). Their morphology was observed by microscopy (A); cell death, GFP expression, and viral progeny production were determined (B) (mean \pm SD of triplicates, representative of three experiments).

* $p < 0.05$ and *** $p < 0.001$ (two-tailed unpaired t test).

(C) Bax/Bak WT and double-KO MEFs were treated with vehicle (DMSO) or with the caspase inhibitor Q-VD-OPH (10 μ M), and the expression of ISGs was measured 48 hr later by RT-PCR (mean \pm SD of triplicates, representative of three experiments).

(D) Expression of IFN β mRNA by Casp9 WT and KO immortalized MEFs after 6 hr of treatment with vehicle (DMSO) or with the Bcl-2 inhibitor ABT-737 (10 μ M) (mean \pm SD of triplicates, representative of three independent experiments).

(E) Expression of IFN β mRNA by WT primary MEFs at the indicated time points after stimulation with vehicle (DMSO), Bcl-2 inhibitor (ABT-737, 10 μ M), caspase inhibitor (Q-VD-OPH, 10 μ M) or both inhibitors (mean \pm SD of duplicates, representative of three independent experiments).

(F) Expression of IFN β mRNA by Bax/Bak WT and double-KO immortalized MEFs after 6 hr of treatment with vehicle or ABT-737 + Q-VD-OPH (mean \pm SD of triplicates, representative of three independent experiments).

(G) Expression of IFN β mRNA by human PBMCs after 6 hr of treatment with vehicle or ABT-737 + Q-VD-OPH (n = 4 healthy donors, results combined from 2 independent experiments; p value calculated by two-tailed unpaired Student's t test).

* $p < 0.05$; ns, not significant; pairwise comparisons following two-way ANOVA (C, D, and F). See also Figure S6.

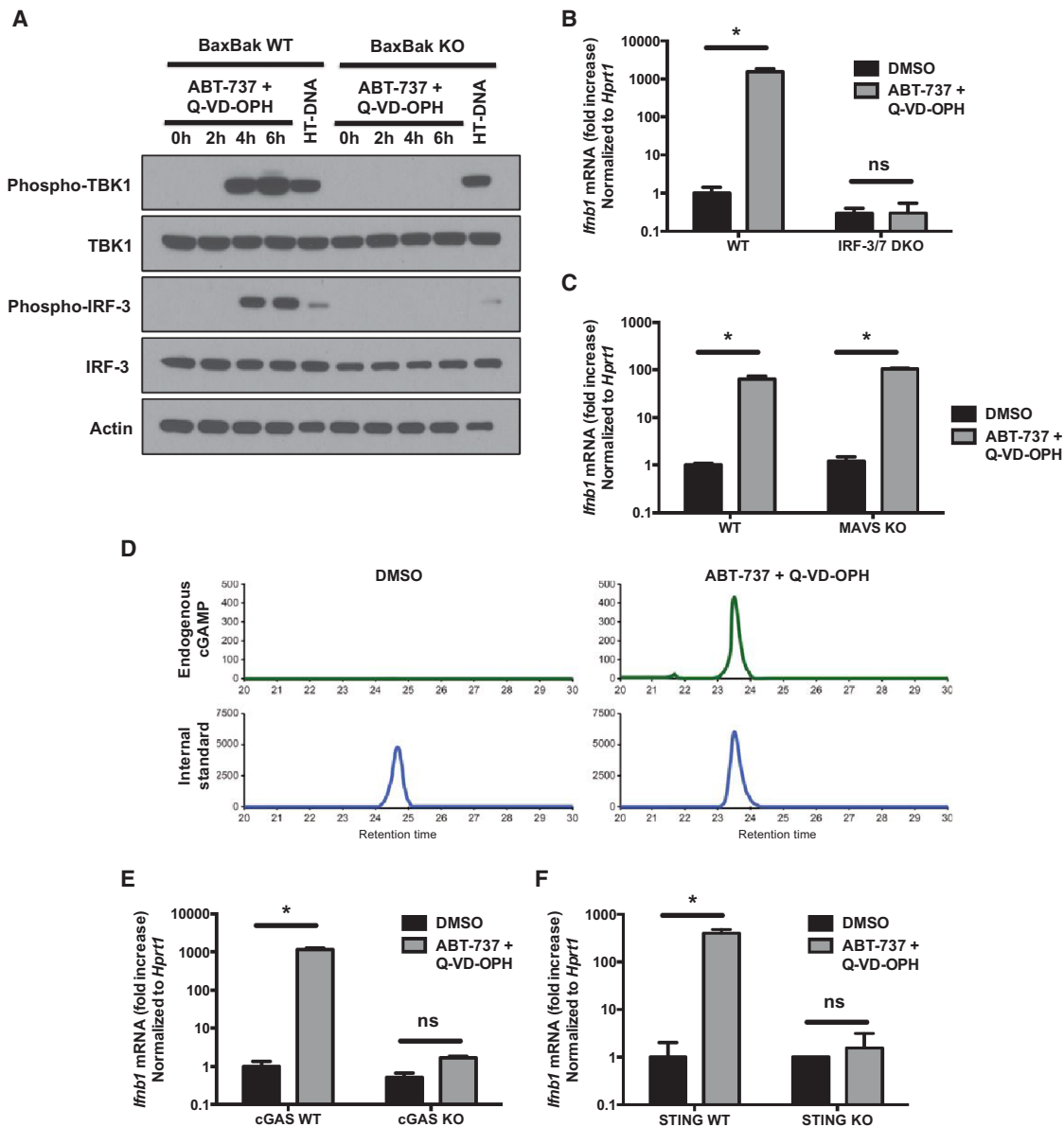


Figure 5. Activation of the cGAS/STING Pathway of IFN Induction

(A) Western blot analysis of the phosphorylation of TBK1 and IRF-3 in Bax/Bak WT and double-KO cells treated with combined Bcl-2/caspase inhibitors (ABT-737 + Q-VD-OPH, 10 μ M each), or transfected with HT-DNA as a positive control (3 μ g/ml, 3 hr). Result representative of three independent experiments.

(B and C) Expression of IFN β mRNA by IRF-3/7 double-KO (B), MAVS KO (C), and control WT primary MEFs after 6 hr of treatment with vehicle (DMSO) or with ABT-737 + Q-VD-OPH (mean \pm SD of triplicates, representative of two experiments).

(D) cGAMP measurement in cell extracts of WT primary MEFs stimulated for 4 hr with vehicle (DMSO) or ABT-737 + Q-VD-OPH. Result representative of two independent experiments.

(E and F) Expression of IFN β mRNA by cGAS WT and KO bone-marrow-derived macrophages (E) and by STING WT and KO primary MEFs (F) after 6 hr of treatment with vehicle or ABT-737 + Q-VD-OPH (mean \pm SD of three or two replicates, respectively).

* $p < 0.05$; ns, not significant; pairwise comparisons following two-way ANOVA (B, C, E, and F). See also Figure S7.

pathway is induced upon recognition of double-stranded DNA by cGAS (Cai et al., 2014). cGAS then acquires its enzymatic activity and synthesizes cGAMP, a dinucleotide that binds to and activates STING. The treatment with ABT-737+Q-VD-OPH resulted in detectable amounts of cGAMP in cell extracts, indicative of cGAS activity (Figure 5D). Furthermore, cGAS deficiency

or STING deficiency completely prevented the IFN β response to ABT-737 + Q-VD-OPH (Figures 5E and 5F). In contrast, TLR signaling was dispensable for the response to ABT-737 + Q-VD-OPH (Figure S7). These results unequivocally identify cGAS/STING as the pathway through which type I IFNs are induced by Bcl-2/caspase coinhibition.

Next, we determined whether the cGAS/STING/TBK1/IRF3 pathway was also responsible for the increased steady-state IFN response in caspase-deficient cells. Interestingly, we observed constitutive phosphorylation of TBK1 in Casp9 KO cells, and this phosphorylation could be further induced by treatment with ABT-737 or transfected HT-DNA (Figure 6A). That observation prompted us to test whether other components of the pathway are constitutively active in the absence of functional caspases. Upon stimulation, IRF-3 is ubiquitinated and degraded through a process of negative feedback loop (Saitoh et al., 2006). We observed lower levels of total IRF-3 protein in Casp9 KO cells, which is suggestive of constitutive activation followed by degradation of IRF-3 (Figure 6A). Similarly, activated STING is phosphorylated and marked for lysosomal degradation (Konno et al., 2013). Again suggestive of constitutive activation, STING protein abundance was reduced in Casp9 KO cells, and the level of STING could be restored to WT levels after treatment with chloroquine, a potent inhibitor of lysosomal acidification (Figure 6B). These observations suggest that the STING pathway is constitutively active in Casp9 KO cells and likely contributes to the IFN-dependent induction of ISG expression. Furthermore, the constitutive activation of STING suggests that the inhibitory activity of caspases acts upstream of STING activation.

We next confirmed the role of the STING pathway at the genetic level. Similarly to the response to ABT-737+Q-VD-OPH, the constitutive expression of ISGs in Casp9 KO cells or in cells treated for 48 hr with the caspase inhibitor Q-VD-OPH was entirely abrogated in cells deficient for IRF-3/7, STING, or cGAS (Figures 6C–6E). In contrast, the RNA recognition pathway was not involved, as the expression of ISGs was not affected by MAVS deficiency (Figure 6F).

Taken together, these results demonstrate that the putative mitochondrial factor released after MOMP and that induces type I IFN in the absence of caspases is a ligand for the cytosolic DNA sensor cGAS. These observations demonstrate the existence of a regulated mechanism of activation of the cGAS pathway by an endogenous ligand. This ligand is sequestered in mitochondria in healthy cells, it is released in a Bax/Bak-dependent manner in dying cells, and its function is intrinsically regulated in a caspase-dependent manner.

mtDNA-Dependent Expression of Type I IFNs

We hypothesized that the mitochondrial ligand recognized by the DNA sensor cGAS could be mitochondrial DNA (mtDNA). To test this possibility, we used a well-established protocol of ethidium bromide (EtdBr)-mediated depletion of mtDNA (Hashiguchi and Zhang-Akiyama, 2009). The addition of low concentrations of EtdBr (150–450 ng/ml) to the culture medium results in the intercalation of EtdBr into mtDNA and prevents its replication, but it does not affect the replication of genomic DNA. This treatment induced an ~10-fold reduction in mtDNA (Figure 7A). When we treated mtDNA-depleted cells with both Bcl-2 and caspase inhibitors, the expression of IFN β was strongly inhibited compared to control cells (Figures 7B and 7C). The phosphorylation of TBK1 and IRF-3 in response to the combined inhibitors was also abolished in mtDNA-depleted cells (Figure 7D). These results implicate mtDNA as the major inducer of type I IFN in this system. In contrast, the response to transfected HT-DNA was

not affected by the EtdBr treatment (Figures 7B and 7C), showing that the cGAS/STING pathway remained functional. The response to other stimuli, such as transfected poly(I:C) and lipopolysaccharide (LPS), was also maintained after EtdBr treatment (Figure 7C), showing that other platforms of innate immune activation are unaffected by EtdBr treatment and mtDNA depletion. The expression of IFN β by Casp9 KO cells in response to Bcl-2 inhibition and the constitutive phosphorylation of TBK1 in unstimulated Casp9 KO cells were also abrogated after depletion of mtDNA with EtdBr (Figures 7E and 7F).

Similarly to the response to ABT-737+Q-VD-OPH, the constitutive expression of ISGs in Casp9 KO cells was reversed by the treatment with EtdBr (Figure 7G), again implicating a role for mtDNA. We confirmed this result with an independent protocol of mtDNA depletion: dideoxycytidine (ddC) is an inhibitor of mitochondrial DNA polymerase γ and does not affect the function of nuclear DNA polymerases (Kaguni, 2004). ddC efficiently depleted mtDNA and reduced the expression of ISGs (Figure 7H), similar to the EtdBr treatment.

Together, these results show that mtDNA is an endogenous ligand that is released from mitochondria via Bax/Bak and that induces type I IFN expression through the cGAS/STING pathway. Caspases play a crucial role in preventing this cell-intrinsic immune response, thus maintaining the immunologically silent nature of mitochondria-dependent apoptotic cell death (Figure 7I).

DISCUSSION

The physiological role of regulated cell death is to maintain homeostasis, and dysregulated cell death can result in cancer, autoimmune and inflammatory disorders, immunodeficiency, or neurodegeneration. The highly regulated process of caspase-dependent apoptosis is unique in its capacity to induce a noninflammatory type of cell death (Martin et al., 2012; Tait et al., 2014). In contrast, caspase-independent cell death generally induces an inflammatory response through release of molecules, termed DAMPs, into the extracellular environment. These DAMPs contribute to the recruitment and activation of inflammatory cells of the immune system such as granulocytes and monocytes/macrophages. Here, we identify a mechanism by which dying cells expose an intracellular DAMP that activates a cell-intrinsic innate immune response. This type of cell-intrinsic immune activation in dying cells could occur while the physical integrity of the plasma membrane is still intact or even in cells that will eventually recover and will not undergo death.

Another singular aspect of this process is the dual role played by mitochondria. Mitochondrial membrane permeabilization is a point of no return in the decision to initiate cell suicide. Bax/Bak-dependent apoptosis is generally considered as a noninflammatory type of cell death. However, our results show that Bax and Bak contribute actively to the induction of the IFN response. The concomitant activation of caspases is required to maintain this type of cell death immunologically silent. This mechanism likely provides the cell with an additional level of control over the decision of whether to die with or without alerting the immune system. One physiological situation in which Bax/Bak-dependent induction of type I IFNs could occur is in the context of infection by viruses that express caspase inhibitors (Best,

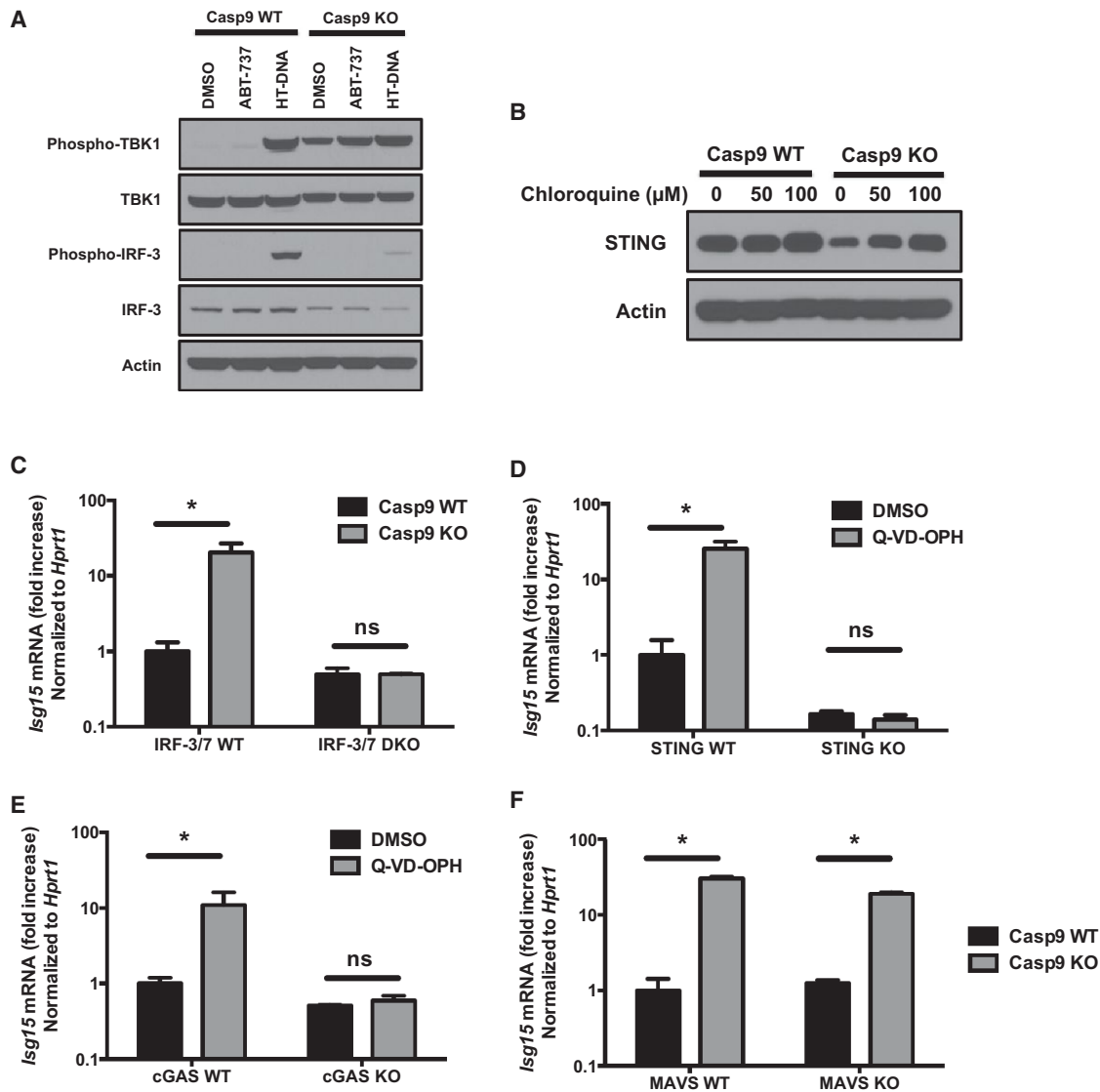


Figure 6. cGAS/STING-Dependent Constitutive ISG Expression in the Absence of Active Caspases

(A) Western blot analysis of the phosphorylation of TBK1 and IRF-3 in Casp9 WT and KO cells treated for 6 hr with vehicle (DMSO), with the Bcl-2 inhibitor ABT-737 (10 μ M), or transfected with HT-DNA as a positive control (3 μ g/ml, 3 hr). Result is representative of three independent experiments.

(B) Western blot analysis of STING in Casp9 WT and KO cells treated for 16 hr with the indicated concentrations of chloroquine.

(C) Caspase-9 KO mice were crossed with IRF-3/7 DKO, and the expression of ISG15 in embryo heads was measured by RT-PCR. Results shown are mean \pm SD of three embryos for each genotype.

(D and E) STING WT and KO primary MEFs (D) or cGAS WT and KO bone-marrow-derived macrophages (E) were treated with vehicle (DMSO) or with the caspase inhibitor Q-VD-OPH (10 μ M) and ISG expression was measured 48 hr later by RT-PCR (mean \pm SD of triplicates, representative of two independent experiments).

(F) Caspase-9 KO mice were crossed with MAVS KO, and the expression of ISGs in embryo heads was measured by RT-PCR. Results shown are mean \pm SD of two embryos for each genotype.

* $p < 0.05$; ns, not significant; pairwise comparisons following two-way ANOVA (C–F).

2008; Callus and Vaux, 2007; Tait et al., 2014). Those virally encoded caspase inhibitors represent an evolutionary response of viruses to host antiviral defenses (i.e., the suicide of infected cells) (Best, 2008; Callus and Vaux, 2007). Our model suggests that sensing caspase inhibition could be a mechanism by which host cells trigger an antiviral response independently of the physical sensing of viral nucleic acids.

Three important questions remain to be elucidated:

- (1) What is the nature of the mtDNA involved, and how does it come into contact with cGAS? Given the size of the Bax/Bak pore, it is likely that small fragments of mtDNA, rather than entire copies of mitochondrial genome, are released. Because such fragments are experimentally difficult to

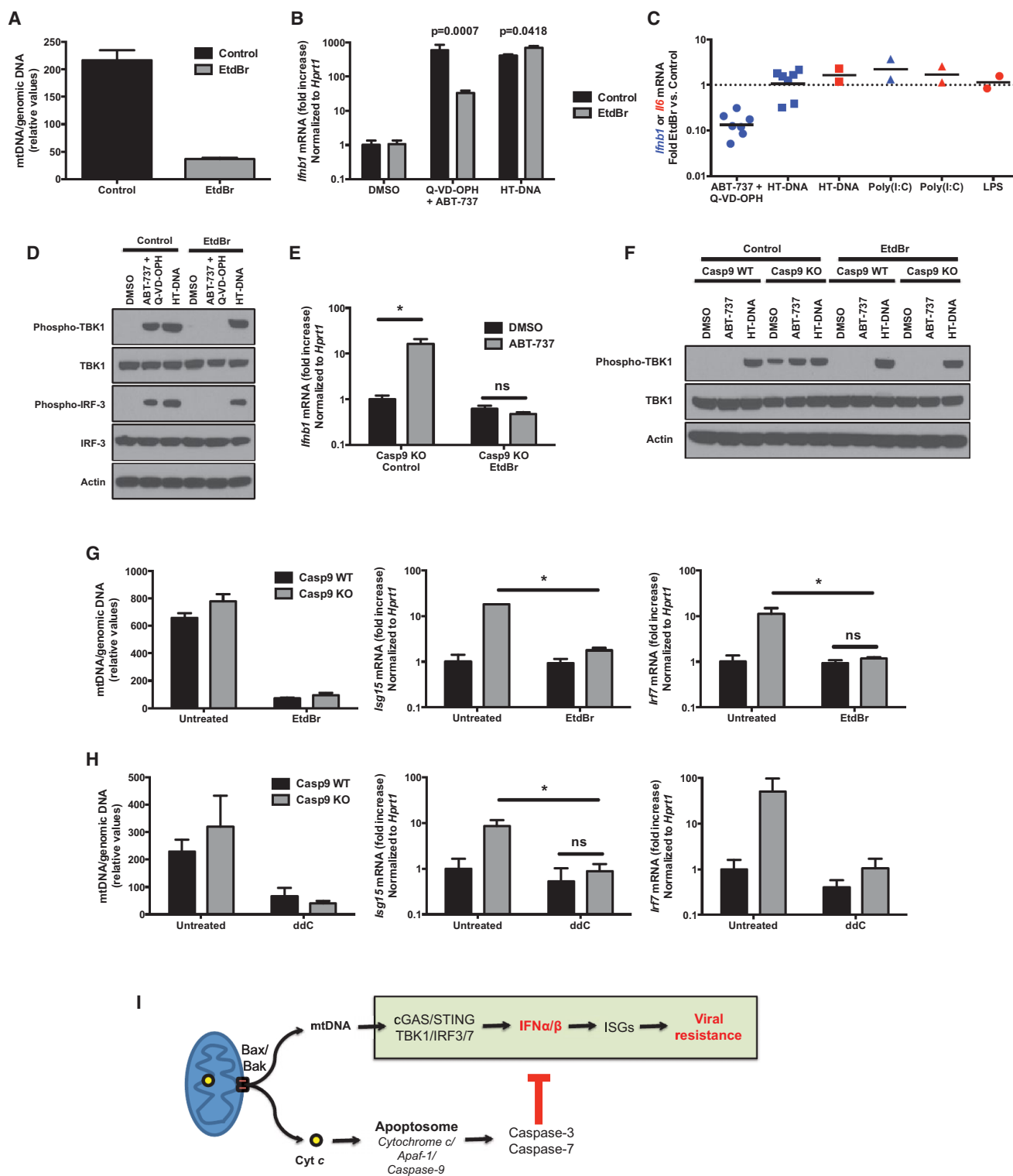


Figure 7. mtDNA Mediates the Induction of Type I IFN Expression

(A) Ratio of mitochondrial DNA (*dloop*) to genomic DNA (*Tert*) measured by RT-PCR on total extracts of WT immortalized MEFs treated for 4 days in EtdBr (150 ng/ml) and then maintained in culture for 16 hr without treatment (mean \pm SD of duplicates, representative of at least five independent experiments).

(B) WT immortalized MEFs treated with EtdBr (150 ng/ml) as in (A) were stimulated with combined Bcl-2/caspase inhibitors (ABT-737 + Q-VD-OPH, 10 μ M each) or transfected HT-DNA (3 μ g/ml) for 6 hr, and the expression of IFN β mRNA was measured by real-time RT-PCR (mean \pm SD of duplicates). p values calculated by two-tailed unpaired Student's t test.

(legend continued on next page)

detect and to quantify reliably, further studies are needed to investigate this possibility. Studying the process of mtDNA release is complicated by the technical challenges of isolating cytosolic and mitochondrial fractions while maintaining the absolute integrity of mitochondria (without release of mtDNA) and of nucleic acids. Furthermore, our knowledge of the physiological mechanisms of mtDNA turnover and degradation is still incomplete (Clay Montier et al., 2009).

- (2) How do caspases prevent mtDNA-dependent activation of the cGAS pathway? Our results suggest that caspases act upstream of STING activation. cGAS or a regulator of cGAS could be targets for caspase-dependent cleavage. Another possibility is that caspase-dependent nucleases (Nagata, 2005) could degrade mtDNA, thus preventing its binding to cGAS. Finally, caspases could affect the release of mtDNA from mitochondria.
- (3) In which cells does this process occur in vivo? Dying cells, or cells that undergo incomplete MOMP and do not die, are the most probable source of Bax/Bak-dependent caspase-regulated type I IFNs. Alternatively, all cells could produce type I IFN at low levels and/or transiently due to leakiness of the Bax/Bak pore in healthy cells.

Answering those questions experimentally will require the development of novel experimental protocols that would allow the simultaneous monitoring of mtDNA release, of cGAS activity, and of caspase activation at the single-cell level.

A surprising observation is that caspase-deficient animals do not develop any symptoms of autoimmune disease, despite the constitutive activation of the type I IFN response. This observation suggests that caspase deficiency could affect the function of other aspects of the immune response. Future studies will determine how caspases contribute to the regulation of adaptive immunity and autoimmunity.

Finally, numerous pharmacological inhibitors of caspases have been developed and are being tested in clinical trials with the aim of preventing tissue damage caused by pathological cell death (Callus and Vaux, 2007). Our results suggest that the consequences of a chronic IFN response induced by these inhibitors should be evaluated. Conversely, we propose that caspase inhibitors could be used to induce an IFN response (i.e., the expression of ISGs) while minimizing the adverse effects caused

by interferon therapies (Vilcek, 2006), as caspase inhibition induces only very low levels of type I IFNs.

It has long been known that apoptosis is an immunologically silent form of cell death, but the molecular basis of this is unknown. We demonstrate a role for mitochondria in the induction of a type I IFN-mediated cell intrinsic immune response in dying cells. Caspases, activated by mitochondria, are required to silence that immune process in apoptotic cells.

EXPERIMENTAL PROCEDURES

Mice

Conditional KO mice with a floxed caspase-9 allele or a floxed caspase-3 allele were generated as described in the Extended Experimental Procedures and illustrated in Figure S1. Caspase-9, caspase-3, caspase-7, Apaf-1, IFNAR1, IRF-3, IRF-7, MAVS, and cGAS (*Mb21d1*^{-/-}) -deficient mice have been reported previously (Honda et al., 2005; Kuida et al., 1996, 1998; Lakhani et al., 2006; Li et al., 2013; Müller et al., 1994; Sato et al., 2000; Sun et al., 2006; Yoshida et al., 1998). All animal experimentations were performed in compliance with Yale Institutional Animal Care and Use Committee protocols.

Cell Cultures

Primary MEFs were generated from caspase-9 KO, IFNAR1 KO, caspase-9/IFNAR1 double KO, caspases-3/-7 double KO, Apaf-1 KO, caspase-9/IRF-3/IRF-7 triple KO, and caspase-9/MAVS double KO and respective littermate control embryos (E16.5–E18.5). All primary MEFs used for experiments were from passage 4 or less. Bax/Bak double-KO and control immortalized MEFs were provided by Dr. C. Thompson (University of Pennsylvania) (Wei et al., 2001), and primary STING KO (*Tmem173*^{-/-}) MEFs were provided by Dr. G. Barber (University of Miami) (Ishikawa et al., 2009). SV40-immortalized Casp9 WT and KO MEFs were reported previously (Masud et al., 2007).

Herring testis DNA (HT-DNA, Sigma-Aldrich, 3 µg/ml) and poly(I:C) (Invivo-gen, 1 µg/ml) were transfected using Lipofectamine 2000 (Invitrogen, 3 µl/ml). IFN α (used at 50 U/ml) and anti-IFN α / β antibodies (used at 300 neutralizing U/ml each) were obtained from Hycult biotech and PBL Assay Science, respectively. Z-VAD-fmk, Boc-D-fmk (EMD Millipore), Q-VD-OPH (MP Biomedicals), and ABT-737 (SantaCruz Biotechnology) were used at 10 µg/ml. Staurosporine and Etoposide were obtained from Sigma-Aldrich and used at 0.01 µM or 10 µM, respectively.

Viral Infections

Mice were infected with EMCV by intraperitoneal injection of 2×10^3 TCID₅₀ of the virus diluted in 100 µl PBS. Mice were sacrificed and hearts were harvested 48 hr later, or survival was monitored for 27 days.

For VSV infection, mice were anesthetized with methoxyflurane (Anafane), and 10^6 PFU of VSV in 50 µl PBS were administered intranasally. The mice were sacrificed 24 hr later. Blood was collected by cardiac puncture and transferred in heparinized tubes. The samples were then centrifuged (2 min at 3,000 rpm), and dilutions of the plasma were used for viral titration.

(C) Fold inhibition by EtdBr pretreatment of the induction of IFN β (blue symbols) or IL-6 (red symbols) mRNA in cells stimulated with ABT-737 + Q-VD-OPH, transfected with HT-DNA, transfected with poly(I:C), or stimulated with LPS. Each dot represents an individual experiment.

(D) Western blot analysis of the phosphorylation of TBK1 and IRF-3 induced by ABT-737 + Q-VD-OPH (10 µM each, 6 hr) or by transfection of HT-DNA (3 µg/ml, 3 hr) in control WT immortalized MEFs or in the same cells pretreated as in (A) with EtdBr (450 ng/ml). Result representative of three independent experiments.

(E) Casp9 KO immortalized MEFs treated or not with EtdBr (450 ng/ml) as in (A) were stimulated with vehicle (DMSO) or the Bcl-2 inhibitor ABT-737 (10 µM) for 6 hr and the expression of IFN β mRNA was measured by real-time RT-PCR (mean \pm SD of duplicates, representative of two independent experiments).

(F) Western blot analysis of the phosphorylation of TBK1 after treatment with vehicle (DMSO) or the Bcl-2 inhibitor ABT-737 (10 µM, 6 hr) or after transfection of HT-DNA (3 µg/ml, 3 hr) in Casp9 WT and KO immortalized MEFs, pretreated or not with EtdBr (450 ng/ml) as in (A). Results are representative of three independent experiments.

(G and H) Casp9 WT and KO primary MEFs were treated for 4 days with EtdBr (150 ng/ml) (G), or immortalized MEFs were treated for 6 days with dideoxycytidine (ddC, 40 µg/ml) (H). The ratio of mitochondrial to genomic DNA was measured by real-time PCR on total extracts (left) and the expression of ISGs was determined by real-time RT-PCR. Results are shown as mean \pm SD of triplicates, representative of three and two independent experiments, respectively.

(I) Schematic model representation of Bax/Bak-dependent, caspase-regulated activation by mtDNA of the cGAS/STING pathway of type I IFN induction.

*p < 0.05; ns, not significant; pairwise comparisons following two-way ANOVA (E, G, and H).

For in vitro infections, cells were plated at a density of 10^5 cells/ml in 6-well plates. The next day, the cells were infected with VSV-GFP, VSV-DsRed, HSV-2, or EMCV (diluted in 500 μ l of DMEM without FBS) for 1 hr with gentle shaking every 15 min. The cells were then washed and incubated in 1 ml of media containing 10% FBS. The dose of virus used for infection and the duration of the incubation are indicated in figure legends.

Cell death was measured by flow cytometry after staining with Annexin V-APC and propidium iodide (BD Biosciences) or by LDH release assay (CytoTox 96 assay, Promega).

Type I IFN Bioassay

To measure type I IFN bioactivity, undiluted culture supernatants were transferred on cultures of the L929-pISRE-Luc reporter cell line (Jiang et al., 2005). The luciferase activity in cell lysates was measured 24 hr later (Dual-Luciferase Reporter Assay, Promega).

Mitochondrial DNA Depletion

Cells were treated with ethidium bromide (150 ng/ml or 450 ng/ml for 4 days; Sigma-Aldrich) or dideoxycytidine (40 μ g/ml for 6 days; Sigma-Aldrich), RNA was extracted, and the expression of ISGs was measured by real-time RT-PCR. For induction of IFN β expression by mtDNA-depleted cells, the cells were cultivated for 4 days in the presence of EtdBr, replated, cultivated overnight in the absence of EtdBr, and then stimulated as indicated. To measure the efficiency of mtDNA depletion, total extracts were prepared by resuspending the cells in NaOH 50 mM, incubation at 95°C for 1 hr, and neutralization by adding 10% volume Tris 1M (pH 7.5). The ratio of mtDNA (*dloop*) versus genomic DNA (*Tert*) was measured by SybrGreen real-time PCR using the following primer pairs:

dloop Forward: AATCTACCATCCTCCGTGAACCC
dloop Reverse: TCAGTTTAGCTACCCCAAGTTTAA
Tert Forward: CTAGCTCATGTGTCAAGACCCTCTT
Tert Reverse: GCCAGCACGTTTCTCTCGTT

Statistical Analysis

The means of two groups were compared using two-tailed unpaired Student's *t* test. When three groups were compared, we used a one-way ANOVA test. When there were two variables, we used a two-way ANOVA test, followed by Games-Howell or Tukey post hoc test to compare pairs of means. Survival curves were compared using Mantel-Cox test.

ACCESSION NUMBERS

The GEO accession number for the RNA sequencing data reported in this paper is GSE63794.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.037>.

AUTHOR CONTRIBUTIONS

A.R., R.J., T.L., M.R.d.Z., and B.Y. performed experiments. A.R., R.J., C.C.D.H., and R.A.F. analyzed results. C.-Y.K., S.A.L., and A.R. generated conditional KO mice. A.P.W., Y.W., T.T., G.S.S., Z.J.C., and A.I. provided scientific advice and reagents. A.R. and R.A.F. conceived the project and wrote the manuscript. R.A.F. supervised the research.

ACKNOWLEDGMENTS

We thank C. Thompson, G. Barber, T. Mak, R. Medzhitov, B. Beutler, and R. Verma for providing reagents; A. Ferrandino, L. Evangelisti, J. Stein, and C. Hughes for ES cell work; F. Sutterwala and N. Palm for comments on the manuscript; J. Alderman for managerial support; and C. Lieber for manuscript

submission. This work was funded by NIAID AI082030, DOD W81XWH-11-1-0745, and Blavatnik Family Foundation M157176 (to R.A.F.); by NIH R01-AI093967 and Cancer Prevention and Research Institute of Texas RP120718-P3 (to Z.J.C.); NIH R01-AG047632 and the United Mitochondrial Disease Foundation (to G.S.S.); by a Grant-In-Aid for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (to T.T.); and by a postdoctoral fellowship PF-13-035-01-DMC from the American Cancer Society (to A.P.W.). C.C.D.H. is a Howard Hughes Medical Institute International Student Research fellow. Z.J.C., A.I., and R.A.F. are Investigators of the Howard Hughes Medical Institute.

Received: June 10, 2014

Revised: October 3, 2014

Accepted: November 10, 2014

Published: December 18, 2014

REFERENCES

- Best, S.M. (2008). Viral subversion of apoptotic enzymes: escape from death row. *Annu. Rev. Microbiol.* 62, 171–192.
- Cai, X., Chiu, Y.H., and Chen, Z.J. (2014). The cGAS-cGAMP-STING pathway of cytosolic DNA sensing and signaling. *Mol. Cell* 54, 289–296.
- Callus, B.A., and Vaux, D.L. (2007). Caspase inhibitors: viral, cellular and chemical. *Cell Death Differ.* 14, 73–78.
- Chipuk, J.E., and Green, D.R. (2005). Do inducers of apoptosis trigger caspase-independent cell death? *Nat. Rev. Mol. Cell Biol.* 6, 268–275.
- Chipuk, J.E., Moldoveanu, T., Llambi, F., Parsons, M.J., and Green, D.R. (2010). The BCL-2 family reunion. *Mol. Cell* 37, 299–310.
- Clay Montier, L.L., Deng, J.J., and Bai, Y. (2009). Number matters: control of mammalian mitochondrial DNA copy number. *J. Genet. Genomics* 36, 125–131.
- Fuchs, Y., and Steller, H. (2011). Programmed cell death in animal development and disease. *Cell* 147, 742–758.
- Gough, D.J., Messina, N.L., Clarke, C.J., Johnstone, R.W., and Levy, D.E. (2012). Constitutive type I interferon modulates homeostatic balance through tonic signaling. *Immunity* 36, 166–174.
- Hakem, R., Hakem, A., Duncan, G.S., Henderson, J.T., Woo, M., Soengas, M.S., Elia, A., de la Pompa, J.L., Kagi, D., Khoo, W., et al. (1998). Differential requirement for caspase 9 in apoptotic pathways in vivo. *Cell* 94, 339–352.
- Hashiguchi, K., and Zhang-Akiyama, Q.M. (2009). Establishment of human cell lines lacking mitochondrial DNA. *Methods Mol. Biol.* 554, 383–391.
- Honda, K., Yanai, H., Negishi, H., Asagiri, M., Sato, M., Mizutani, T., Shimada, N., Ohba, Y., Takaoka, A., Yoshida, N., and Taniguchi, T. (2005). IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* 434, 772–777.
- Ishikawa, H., Ma, Z., and Barber, G.N. (2009). STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity. *Nature* 461, 788–792.
- Jiang, X., and Wang, X. (2004). Cytochrome C-mediated apoptosis. *Annu. Rev. Biochem.* 73, 87–106.
- Jiang, Z., Georgel, P., Du, X., Shamel, L., Sovath, S., Mudd, S., Huber, M., Kalis, C., Keck, S., Galanos, C., et al. (2005). CD14 is required for MyD88-independent LPS signaling. *Nat. Immunol.* 6, 565–570.
- Kaguni, L.S. (2004). DNA polymerase gamma, the mitochondrial replicase. *Annu. Rev. Biochem.* 73, 293–320.
- Kawai, T., and Akira, S. (2011). Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* 34, 637–650.
- Koni, P.A., Joshi, S.K., Temann, U.A., Olson, D., Burkly, L., and Flavell, R.A. (2001). Conditional vascular cell adhesion molecule 1 deletion in mice: impaired lymphocyte migration to bone marrow. *J. Exp. Med.* 193, 741–754.
- Konno, H., Konno, K., and Barber, G.N. (2013). Cyclic dinucleotides trigger ULK1 (ATG1) phosphorylation of STING to prevent sustained innate immune signaling. *Cell* 155, 688–698.

- Kroemer, G., Galluzzi, L., Kepp, O., and Zitvogel, L. (2013). Immunogenic cell death in cancer therapy. *Annu. Rev. Immunol.* 31, 51–72.
- Kroemer, G., Galluzzi, L., Vandenabeele, P., Abrams, J., Alnemri, E.S., Baehrecke, E.H., Blagosklonny, M.V., El-Deiry, W.S., Golstein, P., Green, D.R., et al.; Nomenclature Committee on Cell Death 2009 (2009). Classification of cell death: recommendations of the Nomenclature Committee on Cell Death 2009. *Cell Death Differ.* 16, 3–11.
- Kuida, K., Haydar, T.F., Kuan, C.Y., Gu, Y., Taya, C., Karasuyama, H., Su, M.S., Rakic, P., and Flavell, R.A. (1998). Reduced apoptosis and cytochrome c-mediated caspase activation in mice lacking caspase 9. *Cell* 94, 325–337.
- Kuida, K., Zheng, T.S., Na, S., Kuan, C., Yang, D., Karasuyama, H., Rakic, P., and Flavell, R.A. (1996). Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. *Nature* 384, 368–372.
- Kumar, S. (2007). Caspase function in programmed cell death. *Cell Death Differ.* 14, 32–43.
- Lakhani, S.A., Masud, A., Kuida, K., Porter, G.A., Jr., Booth, C.J., Mehal, W.Z., Inayat, I., and Flavell, R.A. (2006). Caspases 3 and 7: key mediators of mitochondrial events of apoptosis. *Science* 311, 847–851.
- Li, X.D., Wu, J., Gao, D., Wang, H., Sun, L., and Chen, Z.J. (2013). Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects. *Science* 341, 1390–1394.
- Loo, Y.M., and Gale, M., Jr. (2011). Immune signaling by RIG-I-like receptors. *Immunity* 34, 680–692.
- Martin, S.J., Henry, C.M., and Cullen, S.P. (2012). A perspective on mammalian caspases as positive and negative regulators of inflammation. *Mol. Cell* 46, 387–397.
- Masud, A., Mohapatra, A., Lakhani, S.A., Ferrandino, A., Hakem, R., and Flavell, R.A. (2007). Endoplasmic reticulum stress-induced death of mouse embryonic fibroblasts requires the intrinsic pathway of apoptosis. *J. Biol. Chem.* 282, 14132–14139.
- Müller, U., Steinhoff, U., Reis, L.F., Hemmi, S., Pavlovic, J., Zinkernagel, R.M., and Aguet, M. (1994). Functional role of type I and type II interferons in antiviral defense. *Science* 264, 1918–1921.
- Nagata, S. (2005). DNA degradation in development and programmed cell death. *Annu. Rev. Immunol.* 23, 853–875.
- Oltersdorf, T., Elmore, S.W., Shoemaker, A.R., Armstrong, R.C., Augeri, D.J., Belli, B.A., Bruncko, M., Deckwerth, T.L., Dinges, J., Hajduk, P.J., et al. (2005). An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature* 435, 677–681.
- Riedl, S.J., and Salvesen, G.S. (2007). The apoptosome: signalling platform of cell death. *Nat. Rev. Mol. Cell Biol.* 8, 405–413.
- Saitoh, T., Tun-Kyi, A., Ryo, A., Yamamoto, M., Finn, G., Fujita, T., Akira, S., Yamamoto, N., Lu, K.P., and Yamaoka, S. (2006). Negative regulation of interferon-regulatory factor 3-dependent innate antiviral response by the prolyl isomerase Pin1. *Nat. Immunol.* 7, 598–605.
- Sato, M., Suemori, H., Hata, N., Asagiri, M., Ogasawara, K., Nakao, K., Nakaya, T., Katsuki, M., Noguchi, S., Tanaka, N., and Taniguchi, T. (2000). Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN- α /beta gene induction. *Immunity* 13, 539–548.
- Schneider, W.M., Chevillotte, M.D., and Rice, C.M. (2014). Interferon-stimulated genes: a complex web of host defenses. *Annu. Rev. Immunol.* 32, 513–545.
- Stetson, D.B. (2009). Connections between antiviral defense and autoimmunity. *Curr. Opin. Immunol.* 21, 244–250.
- Stetson, D.B., and Medzhitov, R. (2006). Type I interferons in host defense. *Immunity* 25, 373–381.
- Sun, Q., Sun, L., Liu, H.H., Chen, X., Seth, R.B., Forman, J., and Chen, Z.J. (2006). The specific and essential role of MAVS in antiviral innate immune responses. *Immunity* 24, 633–642.
- Tait, S.W., and Green, D.R. (2010). Mitochondria and cell death: outer membrane permeabilization and beyond. *Nat. Rev. Mol. Cell Biol.* 11, 621–632.
- Tait, S.W., Ichim, G., and Green, D.R. (2014). Die another way—non-apoptotic mechanisms of cell death. *J. Cell Sci.* 127, 2135–2144.
- Tamura, T., Yanai, H., Savitsky, D., and Taniguchi, T. (2008). The IRF family transcription factors in immunity and oncogenesis. *Annu. Rev. Immunol.* 26, 535–584.
- Taniguchi, T., and Takaoka, A. (2001). A weak signal for strong responses: interferon- α /beta revisited. *Nat. Rev. Mol. Cell Biol.* 2, 378–386.
- Taylor, R.C., Cullen, S.P., and Martin, S.J. (2008). Apoptosis: controlled demolition at the cellular level. *Nat. Rev. Mol. Cell Biol.* 9, 231–241.
- Upton, J.W., and Chan, F.K. (2014). Staying alive: cell death in antiviral immunity. *Mol. Cell* 54, 273–280.
- Vanden Berghe, T., Linkermann, A., Jouan-Lanhuet, S., Walczak, H., and Vandenabeele, P. (2014). Regulated necrosis: the expanding network of non-apoptotic cell death pathways. *Nat. Rev. Mol. Cell Biol.* 15, 135–147.
- Vilcek, J. (2006). Fifty years of interferon research: aiming at a moving target. *Immunity* 25, 343–348.
- Wei, M.C., Zong, W.X., Cheng, E.H., Lindsten, T., Panoutsakopoulou, V., Ross, A.J., Roth, K.A., MacGregor, G.R., Thompson, C.B., and Korsmeyer, S.J. (2001). Proapoptotic BAX and BAK: a requisite gateway to mitochondrial dysfunction and death. *Science* 292, 727–730.
- Yatim, N., and Albert, M.L. (2011). Dying to replicate: the orchestration of the viral life cycle, cell death pathways, and immunity. *Immunity* 35, 478–490.
- Yoshida, H., Kong, Y.Y., Yoshida, R., Elia, A.J., Hakem, A., Hakem, R., Penninger, J.M., and Mak, T.W. (1998). Apaf1 is required for mitochondrial pathways of apoptosis and brain development. *Cell* 94, 739–750.
- Youle, R.J., and Strasser, A. (2008). The BCL-2 protein family: opposing activities that mediate cell death. *Nat. Rev. Mol. Cell Biol.* 9, 47–59.

Dual Proteolytic Pathways Govern Glycolysis and Immune Competence

Wei Lu,^{1,2,14} Yu Zhang,^{3,2,14} David O. McDonald,⁴ Huie Jing,^{3,2} Bernadette Carroll,⁵ Nic Robertson,^{4,6} Qian Zhang,^{3,2} Helen Griffin,⁷ Sharon Sanderson,⁸ Jeremy H. Lakey,⁵ Neil V. Morgan,⁹ Louise N. Reynard,⁴ Lixin Zheng,^{1,2} Heardley M. Murdock,^{2,3} Stuart E. Turvey,¹⁰ Scott J. Hackett,¹¹ Tim Prestidge,¹² Julie M. Hall,⁶ Andrew J. Cant,^{4,6} Helen F. Matthews,^{1,2} Mauro F. Santibanez Koref,⁷ Anna Katharina Simon,^{8,13} Viktor I. Korolchuk,⁵ Michael J. Lenardo,^{1,2} Sophie Hambleton,^{4,6,15,*} and Helen C. Su^{3,2,15,*}

¹Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

²NIAID Clinical Genomics Program, National Institutes of Health, Bethesda, MD 20892, USA

³Laboratory of Host Defenses, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA

⁴Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne NE2 4HH, UK

⁵Institute of Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne NE2 4HH, UK

⁶Great North Children's Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne NE1 4LP, UK

⁷Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne NE1 3BZ, UK

⁸NIHR BRC Translational Immunology Lab, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, UK

⁹Centre for Cardiovascular Sciences, School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, Birmingham B15 2TT, UK

¹⁰Department of Pediatrics, Child & Family Research Institute and BC Children's Hospital, University of British Columbia, Vancouver, BC V5Z 4H4, Canada

¹¹Paediatric Immunology Department, Birmingham Heartlands Hospital, Birmingham B9 5SS, UK

¹²Blood and Cancer Center, Starship Children's Hospital, Auckland 1142, New Zealand

¹³MRC Unit Human Immunology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK

¹⁴Co-first author

¹⁵Co-senior author

*Correspondence: sophie.hambleton@newcastle.ac.uk (S.H.), hsu@niaid.nih.gov (H.C.S.)

<http://dx.doi.org/10.1016/j.cell.2014.12.001>

SUMMARY

Proteasomes and lysosomes constitute the major cellular systems that catabolize proteins to recycle free amino acids for energy and new protein synthesis. Tripeptidyl peptidase II (TPPII) is a large cytosolic proteolytic complex that functions in tandem with the proteasome-ubiquitin protein degradation pathway. We found that autosomal recessive *TPP2* mutations cause recurrent infections, autoimmunity, and neurodevelopmental delay in humans. We show that a major function of TPPII in mammalian cells is to maintain amino acid levels and that TPPII-deficient cells compensate by increasing lysosome number and proteolytic activity. However, the overabundant lysosomes derange cellular metabolism by consuming the key glycolytic enzyme hexokinase-2 through chaperone-mediated autophagy. This reduces glycolysis and impairs the production of effector cytokines, including IFN- γ and IL-1 β . Thus, TPPII controls the balance between intracellular amino acid availability, lysosome number, and glycolysis, which is vital for adaptive and innate immunity and neurodevelopmental health.

INTRODUCTION

Protein degradation occurs continuously within cells. This removes misfolded or damaged proteins and generates free amino acids for protein synthesis or energy production via glutaminolysis (Schutz, 2011). Mammalian cells utilize two principal pathways: proteasomes, which are protein complexes that recognize and degrade ubiquitinated proteins within the cytosol, and lysosomes, which are membrane-bound organelles containing acid hydrolases that are fed substrate by endosomal and autophagic vesicles (Ciechanover, 2005). Evidence suggests that these pathways can cross-compensate to maintain balanced proteolysis and amino acid homeostasis (Korolchuk et al., 2010). In both pathways, proteins are first degraded into long oligopeptides from which N-terminal tripeptides are then trimmed by tripeptidyl peptidases (TPP). These tripeptides are further cleaved by dipeptidyl peptidases and aminopeptidases to generate free amino acids (Tomkinson, 1999).

There are two types of TPP in eukaryotic cells, TPPI and TPPII. TPPI is a lysosomal acid protease, whereas TPPII is a cytosolic protease that forms a giant multisubunit complex acting downstream of proteasomes (Schönege et al., 2012; Tomkinson, 1999). By trimming long oligopeptides, TPPII was thought to be principally important in producing antigenic peptides that bind to major histocompatibility complex (MHC) class I molecules for presentation to CD8 T cells (Reits et al., 2004). However, the development and function of CD8 T cells was

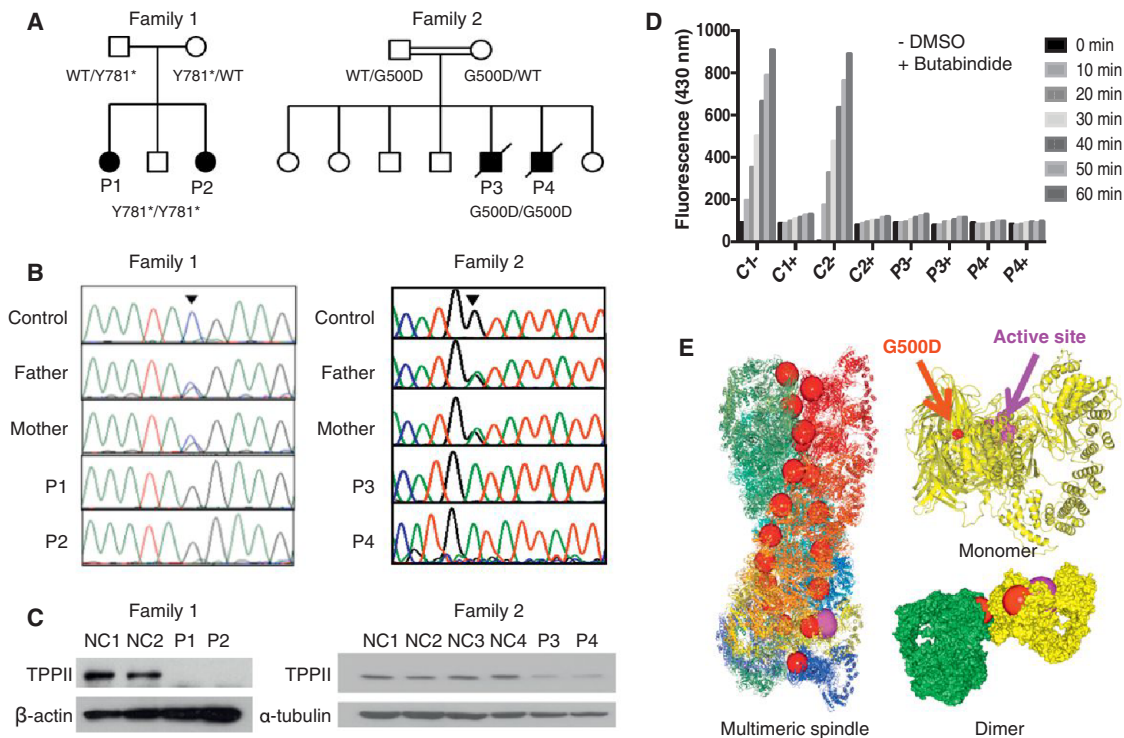


Figure 1. Autosomal Recessive Loss-of-Function Mutations in Human TPP2 Deficiency

(A) Patients' pedigrees.

(B) Sanger sequencing showing the mutations.

(C) Immunoblots for TPP2 in T cells from P1 and P2 (left) or fibroblasts from P3 and P4 (right).

(D) TPP2 enzymatic activity in fibroblast lysates from two healthy controls, P3 and P4, incubated for the indicated minutes without (–) or with (+) added TPP2 inhibitor BUTA.

(E) Structural representations of TPP2 highlighting G500D (red spheres) and the active site (purple spheres). Shown are ribbon representations of multimeric spindle and monomer and surface representation of dimer of yellow and green monomers (Schönege et al., 2012).

Experiments were repeated at least twice for (C) and three times for (D). See also Figure S2.

largely unaffected by genetic deletion of *Tpp2* in mice, even during experimental viral infections (Kawahara et al., 2009). By contrast, other *Tpp2*-deficient mouse strains exhibited either embryonic lethality (McKay et al., 2007) or an immunosenescent phenotype characterized by declining thymic output and progressive loss of CD4 and CD8 T cells (Huai et al., 2008). Thus, the physiological role for TPP2 in proteolysis, amino acid homeostasis, and metabolism in mammals remains obscure. Furthermore, although humans with loss-of-function mutations in *TPP1* develop a lysosomal storage disease called classical late-infantile neuronal ceroid lipofuscinosis (Tomkinson, 1999), whether *TPP2* mutations cause human disease is unknown.

In the immune system, innate and adaptive cells quickly and coordinately respond to invading pathogens and inflammatory signals. The biosynthetic and bioenergetic demands of the responding leukocytes are extreme because of the sudden requirements for cell growth, trafficking, proliferation, and effector functions. To support this burst of anabolic activity, cellular metabolism radically reorients toward aerobic glycolysis (MacIver et al., 2013; Pearce and Pearce, 2013). Although less efficient in generating ATP, glycolysis generates intermediate metabolites that support biosynthetic pathways for effector

functions, including cytokine production (Chang et al., 2013; Shi et al., 2011). It is thus not surprising that metabolic reprogramming is an integral part of leukocyte activation and that a complex regulatory network links nutrient availability with a concerted immune response. Unraveling this complexity is important because of the potential to target metabolic pathways for modulating pathological immune responses. To this end, we have studied patients with a metabolic immunodeficiency caused by *TPP2* mutations.

RESULTS

Human Disease Caused by Loss of TPP2 Activity

We identified four patients from two families, affected by combined immunodeficiency, severe autoimmunity, and developmental delay (Figure 1A, Table 1, and Data S1 available online), with biallelic loss-of-function mutations in *TPP2*. Except for P2, who was diagnosed by screening in early infancy, patients presented in early childhood with recurrent bacterial and viral infections of the respiratory tract and middle ear. All three tested patients showed markedly decreased circulating T, B, and natural killer lymphocytes (Figure S1A), including severely

Table 1. Clinical Features of TPPII-Deficient Patients

Patient	P1	P2	P3	P4
Demographics	10-year-old First Nations female	18-month-old First Nations female	11-year-old British Pakistani male	3-year-old British Pakistani male
Immunodeficiency				
Respiratory Tract	recurrent lower respiratory tract infections (including cytomegalovirus, adenovirus) with bronchiectasis, recurrent otitis media	lower respiratory tract infections (including adenovirus, <i>Aspergillus fumigatus</i>), recurrent otitis media	recurrent lower respiratory tract infections with bronchiectasis, recurrent otitis media complicated by mastoiditis	recurrent lower respiratory tract infections, recurrent otitis media
Other Infections	recurrent orolabial herpes simplex virus type 1	none	pneumococcal sepsis, hemorrhagic varicella, acute hepatitis A, persistent cytomegalovirus	acute hepatitis A, cytomegalovirus
Autoimmunity				
Cytopenias	autoimmune hemolytic anemia, immune thrombocytopenic purpura, neutropenia	neutropenia	autoimmune hemolytic anemia, immune thrombocytopenic purpura	autoimmune hemolytic anemia, immune thrombocytopenic purpura, neutropenia
Other Autoimmunity	central nervous system lupus erythematosus (ANA ⁺) with stroke due to suspected vasculitis	none	none	autoimmune hepatitis (ANA ⁺ SMA ⁺) leading to end-stage liver disease
Developmental delay	yes	yes, assessment ongoing	yes	yes
Outcome	alive	alive, recently underwent hematopoietic stem cell transplantation	died of adenovirus encephalitis after hematopoietic stem cell transplantation	died of complications after orthotopic liver transplantation

ANA, anti-nuclear autoantibodies; SMA, anti-smooth muscle autoantibodies. See also Figure S1 and Data S1.

reduced naive T cells, and hypergammaglobulinemia (Figure S1B and data not shown). Patients had severe, intractable autoimmunity, manifesting as antibody-mediated destruction of red blood cells, platelets, and neutrophils (in all patients, Figure S1C); central nervous system lupus erythematosus with stroke (in P1); and hepatitis with autoantibodies (in P4). All patients had mild to moderate developmental delay in acquiring motor, language, and social skills. We assumed that an autosomal recessive disorder was present, as siblings P1 and P2 originated from a geographically isolated, small indigenous population and sibship P3 and P4 from a consanguineous union. Diagnostic testing excluded known primary immunodeficiency disorders, so we performed whole-exome sequencing (WES) of genomic DNA from P1 and her healthy parent and WES with homozygosity mapping in P3 (Figure S2A). Each patient bore a novel homozygous single nucleotide variant in *TPP2*, for which their parents were heterozygous carriers. Sanger sequencing of *TPP2* confirmed that P1 and P2 were homozygous for the nonsense mutation c.2343C > G, p.Tyr781*, whereas P3 and P4 were homozygous for the missense mutation c.1499G > A, p.Gly500Asp (Figure 1B).

TPPII protein is normally expressed at high levels in lymphocytes (Figure S2B) but was absent from T cells from both P1 and P2, consistent with nonsense-mediated mRNA decay (Figure 1C). In contrast, TPPII protein was expressed by dermal fibroblasts and peripheral blood mononuclear cells (PBMC)

from P3 and P4, although at levels 60%–80% less than controls (Figure 1C and Figure S2B). The residual protein lacked exopeptidase activity sensitive to the TPPII-specific inhibitor butabindide (BUTA, Rose et al., 1996) (Figure 1D). To model the effect of the p.G500D missense mutation, we considered the recently solved quaternary structure of human TPPII (Schönege et al., 2012). TPPII exists as a giant (~6 MDa) spindle-shaped cytoplasmic protein complex made up of two twisted strands composed of stacked dimers that align to form a series of internal catalytic chambers (Figure 1E). Glycine at residue 500, which is conserved across all species (Figure S2C), lies outside of the active site at the hydrophobic interface between strands of stacked dimers (Figure 1E). We predicted that substitution by a charged amino acid would disrupt this “molecular clamp,” thereby impairing multimerization and enzymatic function. In keeping with this hypothesis, native gel electrophoresis revealed loss of high molecular weight complexes of TPPII protein and associated enzymatic activity in lysates from P3 and P4 (Figure S2D).

TPPII Has a Major Role in Amino Acid Homeostasis

During normal cellular homeostasis, cytosolic free amino acids are derived from recycling of intracellular proteins through the proteasomal or lysosomal proteolytic pathways and from extracellular transport (Barnes et al., 1992). Because TPPII is a cytosolic protease that acts downstream of the proteasome, we hypothesized that TPPII deficiency would impair amino

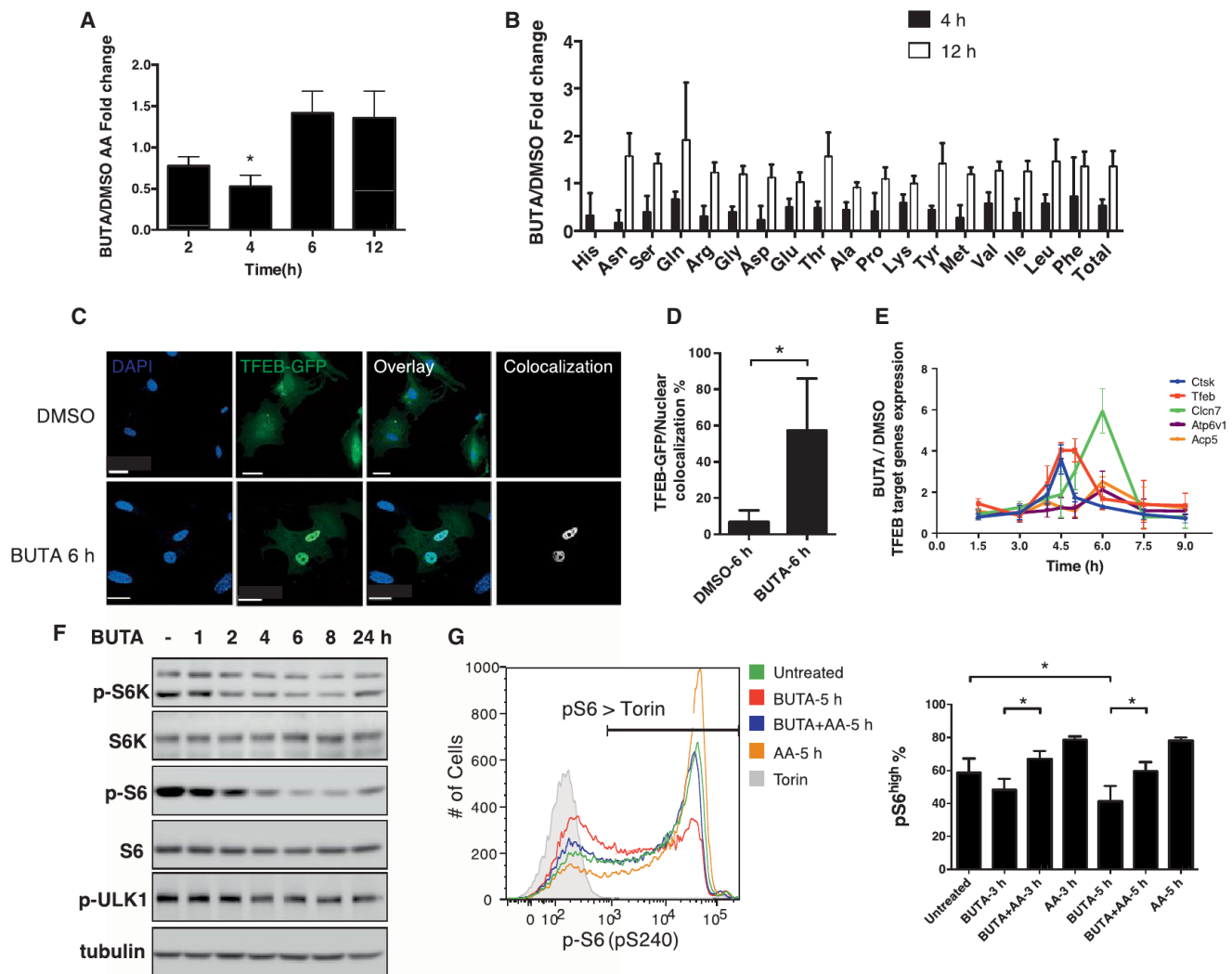


Figure 2. TPPII Regulates Intracellular Amino Acid Homeostasis

(A) Each bar shows the ratio of the quantity of total free amino acids (AA) in wild-type MEF cells treated with BUTA (200 μ M) for the indicated hours (h) to the quantity of total AA in the same cells treated with the dimethyl sulfoxide (DMSO) vehicle, as determined by ultraperformance liquid chromatography (UPLC). (B) Normalized concentrations of intracellular free AA in MEF after 4 or 12 hr of treatment with BUTA or DMSO as in (A). (C) Confocal microscopy showing colocalization (white) of GFP-tagged TFEB (green) and DAPI (blue) in transfected MEF treated with BUTA or DMSO for 6 hr. Scale bars, 30 μ m (BUTA), 40 μ m (DMSO). (D) Quantitation of (C). (E) Normalized expression of TFEB-regulated gene transcripts, measured by quantitative RT-PCR, in MEF treated with BUTA or DMSO for the indicated times. (F) Immunoblot showing kinetics of S6K, S6, and ULK1 phosphorylation in BUTA-treated primary human fibroblasts. (G) Flow cytometric detection of reduced pS6 in Jurkat T cells treated with BUTA and rescue by supraphysiologic AA (mTOR inhibitor torin, negative control); percentages of pS6-positive cells quantified (right). Data in (A), (B), (D), (E), and (G) are represented as mean \pm SD from three independent experiments. * $p < 0.05$, as calculated by one sample t test (A), unpaired two-tailed Student's t test (D), or repeated-measures ANOVA followed by Tukey's test (G). Images in (C), (F), and (G) are representative of three independent experiments. See also Figure S3.

acid recycling. Indeed, total levels of intracellular free amino acids were markedly decreased within 4 hr of TPPII inhibition by BUTA treatment of mouse embryonic fibroblasts (MEF), at values $\sim 20\%$ to $\sim 70\%$ of controls (Figure 2A). This decrease affected all measurable individual amino acids (Figure 2B and data not shown), suggesting no selective effect on amino acids or transporters that recognize structurally related classes of

amino acids (Taylor, 2014). However, by 6–12 hr of BUTA treatment, intracellular free amino acids had returned to normal or higher levels (Figure 2A), similar to chronically TPPII-deficient knockout (KO) MEF cells (Figures S3A and S3B). Thus, a major function of TPPII is to maintain amino acid levels in the cell, and during chronic TPPII inhibition, a compensatory mechanism maintains amino acid homeostasis.

Amino acid deprivation causes the transcription factor TFEB to translocate from the cytosol to the nucleus, where it functions as a master regulator of genes involved in lysosomal biogenesis and autophagy to preserve intracellular amino acid homeostasis (Rocznik-Ferguson et al., 2012; Settembre et al., 2011). We therefore monitored TFEB subcellular localization and target gene transcription upon BUTA treatment. Because commercially available antibodies were unreliable for detecting endogenous TFEB, transfected GFP-tagged TFEB was used. Coincident with the acute decrease in intracellular free amino acids after TPPII inhibition (Figure 2A), we found TFEB translocated into the nuclei of MEF (Figures 2C and 2D) or the human neuroblastoma cell line SH-SY5Y (Figure S3C). Furthermore, in untransfected MEF, BUTA upregulated endogenous TFEB-regulated transcripts, including *Tfeb* itself, within 4–6 hr of treatment before returning to baseline (Figure 2E). We did not detect elevated *TFEB*-regulated transcripts in cells constitutively deficient in TPPII (Figure S3D). Thus, maximal amino acid reduction precedes maximal TFEB induction, which in turn precedes restored amino acid homeostasis.

TFEB nuclear localization is normally inhibited by mTOR when growth and nutritional signals are adequate (Rocznik-Ferguson et al., 2012). We found that basal mTOR activity, as measured by phosphorylation of either S6K or S6, was markedly diminished in fibroblasts after 2–8 hr of BUTA treatment (Figures 2F and S3E) when TFEB activity was highest. Acute TPPII inhibition also decreased mTOR activity in the Jurkat T cell line and PHA-stimulated T cell blasts, which could be overcome by supraphysiological extracellular amino acids (Figures 2G and S3F). Remarkably, mTOR activity recovered when BUTA treatment was extended beyond 18 hr (Figures 2F, S3E, and S3G), and mTOR activity was normal in a THP1 macrophage cell line stably transduced with *TPP2* shRNA (Figure S3H). Such chronically TPPII-inhibited cells nevertheless showed exaggerated sensitivity to partial amino acid starvation (Figure S3G). Together, our results indicate that acute TPPII inhibition caused amino acid depletion, leading to mTOR deactivation, TFEB nuclear translocation, and TFEB-dependent gene induction, thereby adapting the cells to chronic TPPII inhibition.

Compensatory Lysosomal Biogenesis in the Absence of TPPII Activity

Lysosomes are a key regulator of amino acid homeostasis (Bar-Peled and Sabatini, 2014), so we hypothesized that the new metabolic state established in the absence of TPPII activity was achieved through a TFEB-dependent increase in lysosomal activity. We observed increased abundance of lysosomes in T cells from P1 (Figures 3A and 3B), above levels expected from T cell receptor (TCR) stimulation alone (Valdor et al., 2014 and data not shown). Similar effects were observed after BUTA treatment of normal T cells (Figure 3C), human fibroblasts (Figures 3D and 3E), MEF, and various transformed cell lines (293T, A549, SH-SY5Y, HeLa) (Figure S4A) or when TPPII was knocked down in THP1 cells stably expressing *TPP2* shRNA (Figure S4B). The lysosomal markers LAMP1 and cathepsin B were increased in fibroblasts derived from P3 and P4, as well as in control fibroblasts treated with BUTA (Figures 3F, S4C, and data not shown). Measures of lysosomal function, such as

acid phosphatase and β -N-acetyl-glucosaminidase activity, were also commensurately elevated in liver tissues from *Tpp2* KO mice (Figures S4D and S4E). Confirming that this lysosomal expansion was caused by inhibition of cytoplasmic proteolysis, we could mimic the effect by treating MEF with the proteasome inhibitor MG132, which blocks the pathway upstream of TPPII (Figure S4F). Thus, cells of multiple lineages experience impaired cytosolic amino acid recycling when TPPII activity is deficient and then recover through an mTOR- and TFEB-regulated compensatory lysosomal biogenesis that restores amino acid homeostasis.

Lysosomal Overactivity Induced by TPPII Deficiency Impairs Glycolysis

Our patients' neurodevelopmental delay, altered amino acid homeostasis, expanded lysosomal compartment, and, in the eldest patient (P3), an unusual pattern of intracranial calcification involving basal ganglia and subcortical U fibers (Figure S5A) all implied a metabolic disorder. Because metabolic reprogramming is crucial for activated proliferating lymphocytes, we evaluated glycolysis by measuring extracellular acidification rate (ECAR) after providing glucose to TPPII-deficient or -replete cells. We also measured the effect of adding oligomycin (to block oxidative phosphorylation and shunt to glycolysis), as well as 2-deoxy-D-glucose (2-DG, to inhibit glycolysis). TPPII deficiency reduced basal and/or maximal glycolysis in activated T cells from P2 (Figure 4A), as well as BUTA-treated activated CD4 or CD8 T cells from controls (Figures 4B and 4C). We observed similar effects in CD4 T cells isolated from *Tpp2* KO mice (Figure 4D), THP1 cells stably transduced with *TPP2* shRNA (Figure 4E), or in SH-SY5Y cells after BUTA treatment (Figure S5B). By contrast, no defect was evident in unstimulated naive T cells that have yet to switch from mixed fuel oxidative phosphorylation to aerobic glycolysis (Figure S5C). In BUTA-treated activated CD4 T cells, we confirmed the glycolytic defect by showing impaired conversion of radiolabeled glucose to water (Figure 4F). We next evaluated oxidative phosphorylation in activated T cells by measuring oxygen consumption rate (OCR) at baseline, followed by oligomycin (to calculate ATP production), carbonyl cyanide p-trifluoromethoxyphenylhydrazone (FCCP, to determine maximal respiration), and rotenone plus antimycin (to calculate basal respiration) and found that oxidative phosphorylation was unaffected by TPPII deficiency, implying that glutamine utilization was intact (Figures 4G and 4H).

In mammalian cells, glycolysis proceeds through a series of enzymatic reactions that convert glucose to pyruvate and ATP (Figure 5A). We hypothesized that TPPII deficiency reduced glycolytic flux by decreasing glycolytic enzyme activity at one or more of these steps (Figure 5A). Consistent with previous observations that T cell activation upregulates transcripts for glycolytic enzymes (Stentz and Kitabchi, 2004; Wang et al., 2011), immunoblotting revealed that CD3/CD28 stimulation of human naive CD4 T cells from healthy donors induced the expression of hexokinase-2 (HK2), phosphofructokinase (PFKP), pyruvate kinase (PKM2), and, more modestly, that of HK1, aldolase, phosphoglycerate mutase-1 (PGAM1), and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (Figure 5A); addition of IL-2 did not further alter levels (data not shown). Remarkably, treatment

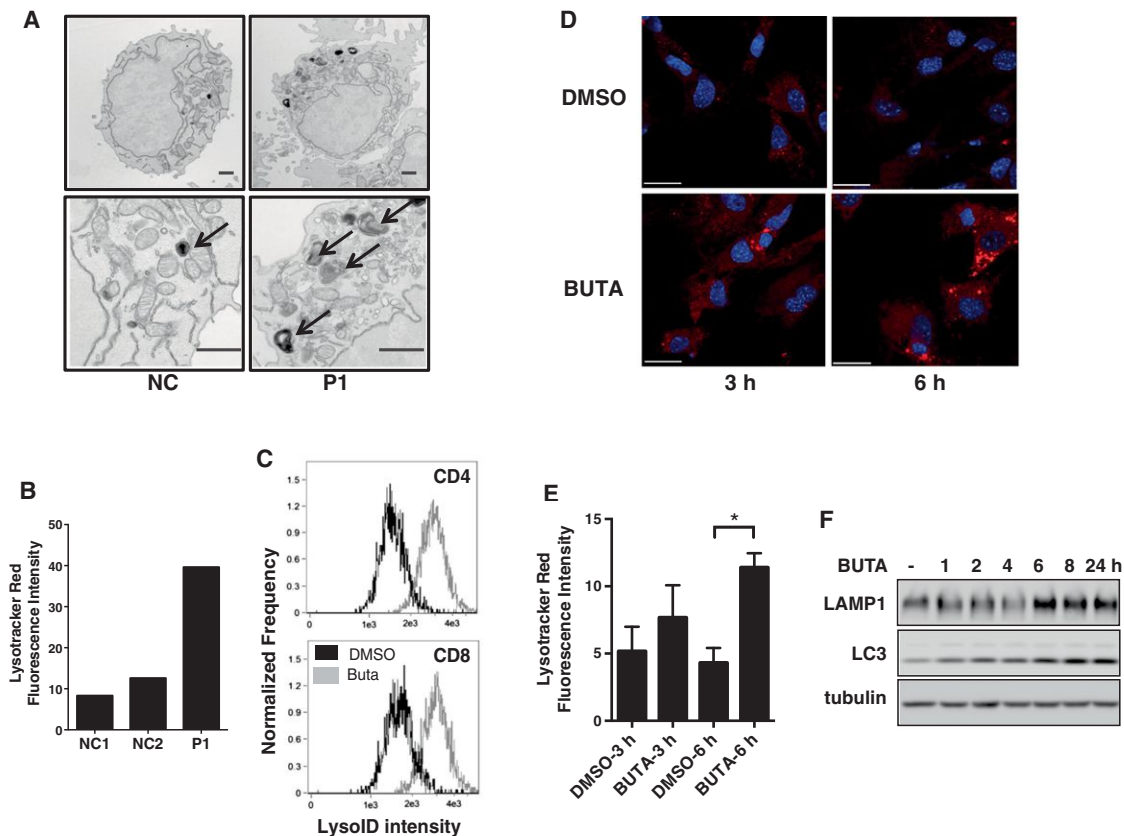


Figure 3. Loss of TPPII Leads to Compensatory Lysosome Biogenesis

(A) Transmission electron micrographs of T cells from P1 and a healthy normal control. Arrows, electron-dense lysosomal structures. Scale bar, 500 nm.

(B) Quantification of Lysotracker Red fluorescence intensity from confocal microscopy of T cells as in (A).

(C) Image stream analysis of lysosome content by LysolD red staining of normal human T cells, activated in the presence of BUTA or DMSO.

(D) Confocal microscopy showing Lysotracker Red staining of lysosomes and DAPI staining of nuclei (blue) after treatment of MEF with DMSO or BUTA for the indicated h. Scale bar, 30 μ m.

(E) Quantification of (D).

(F) Simultaneous immunoblotting of LAMP1 and LC3 following BUTA treatment of human fibroblasts for the indicated h.

All experiments were repeated three times except for (A), which was repeated twice, and show representative images. Data in (E) are represented as mean \pm SD from three independent experiments. * $p < 0.05$, as calculated by unpaired two-tailed Student's *t* test. See also Figure S4.

with BUTA greatly attenuated the induction of HK2 and PKM2 proteins while minimally affecting HK1, PFKP, aldolase, GAPDH, and PGAM1 (Figure 5A and data not shown). By contrast, activated T cells from both P1 and P2 showed markedly decreased HK2 but no consistent decreases in HK1 or PKM2 (Figure 5B). Furthermore, activated T cells from *Tpp2* KO mice showed decreased HK2 and aldolase, but not PGAM1 or PKM2 (Figures 5C and 5D). Similarly, *TPP2* shRNA-transduced THP1 showed decreased HK2 at baseline and after lipopolysaccharide (LPS) stimulation (Figure 5D). Thus, only HK2 was consistently decreased. Because hexokinase performs the rate-limiting first step in glycolysis, this lack of HK2 could explain the profound reduction in glycolytic flux when TPPII was defective, although other glycolytic enzymes may be involved.

To explore the mechanism of HK2 depletion, we first determined that *HK2* mRNA was not decreased in TPPII-defective cells (Figure S5E). Next, we found that HK2 protein could colocalize with lysosomes (Figure S5F). Because TPPII defi-

ciency caused lysosomal expansion (Figures 3A–3E, S4A, and S4B) and lysosomal expansion induced by forced TFEB overexpression decreased HK2 protein levels (Figure S5G), we postulated that HK2 was targeted for lysosomal degradation. Indeed, we observed that treating *Tpp2* KO T cells with lysosomal inhibitors restored HK2 levels (Figures 5C and 5H). Moreover, when protein synthesis was blocked by cycloheximide, HK2 underwent more rapid protein turnover than other glycolytic enzymes (Figure 5E). The glycolytic enzymes aldolase, GAPDH, PGAM1, and PKM2 undergo chaperone-mediated autophagy (CMA), whereby the chaperone HSC70 mediates entry through the lysosomal LAMP2A receptor of substrates containing a KFERQ-like lysosomal degradation targeting peptide motif (Dice, 1990; Kaushik et al., 2011). PKM2 and GAPDH each have one experimentally validated motif, whereas HK2 has two such motifs (Figure 5F). To test their functionality, we overexpressed Myc-tagged HK2 fragments containing either the central QLLEVK motif (residues

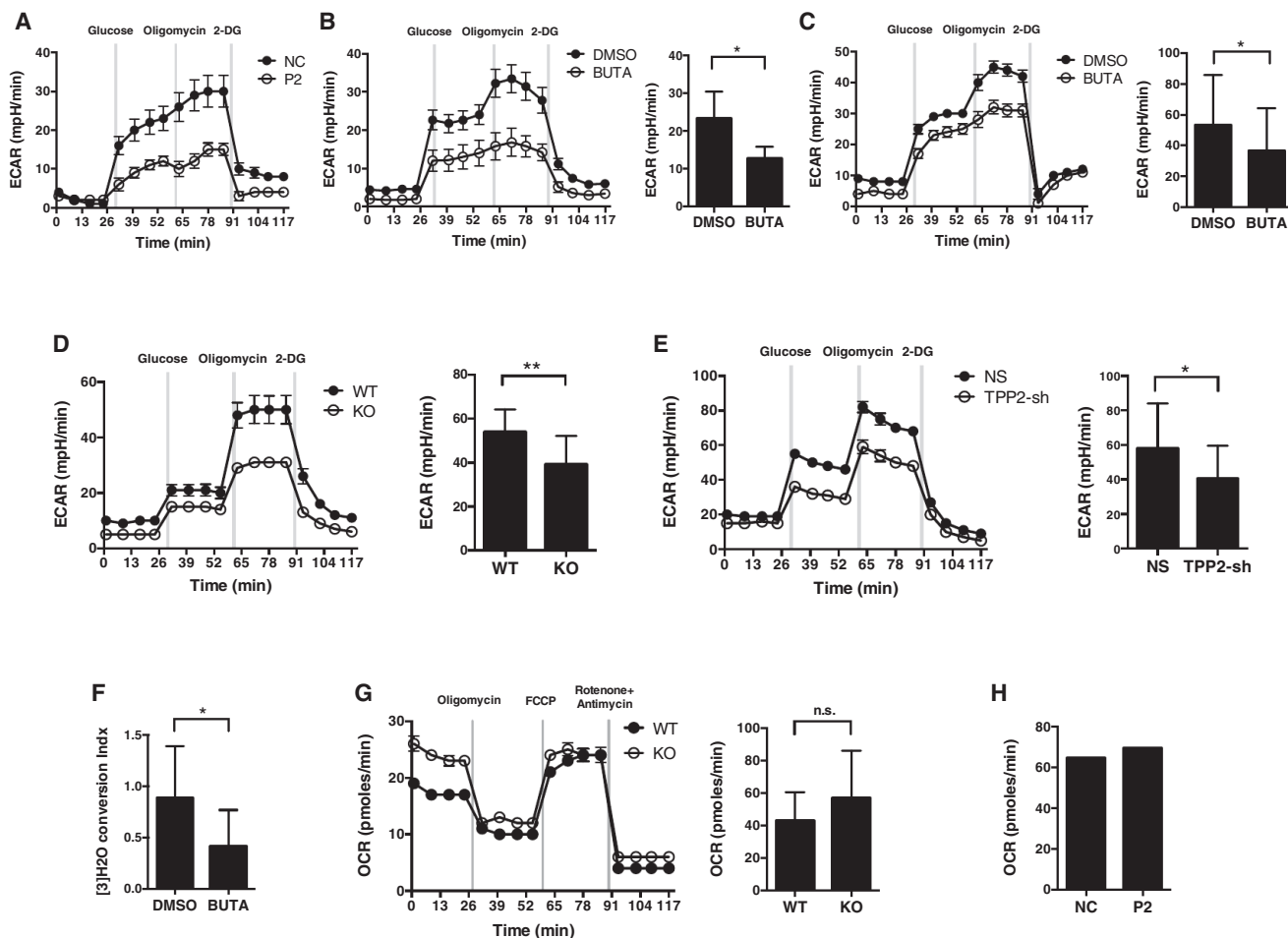


Figure 4. Loss of TPPII Activity Impairs Glycolysis

(A–E) Glycolytic flux, measured as ECAR of cells at baseline and after adding glucose, oligomycin, and 2-DG and/or quantitation by total area under the curve (AUC) of histograms.

(A) Activated T cells from P2 or healthy control.

(B) Human naive CD4 T cells, activated for 3 days in the presence of BUTA (200 μ M) or DMSO.

(C) Human naive CD8 T cells stimulated for 4 days as in (B).

(D) Naive CD4 T cells from *Tpp2* KO or WT mice, activated for 3 days.

(E) THP1 cells, stably transduced with *TPP2* shRNA or NS shRNA.

(F) Glycolysis as measured by [5-³H]-glucose conversion to [5-³H]-water in cells stimulated as in (B).

(G and H) Oxidative phosphorylation as measured by OCR at baseline and after adding oligomycin, FCCP, and rotenone plus antimycin, and/or quantitation by AUC.

(G) Naive CD4 T cells from *Tpp2* KO or WT mice, stimulated as in (D).

(H) T cells from P2 or control as in (A).

For representative histograms, data shows mean \pm SD of six replicate wells. Data showing AUC represent mean \pm SD from at least three independent experiments, except for (H), which was from one experiment. * $p < 0.05$, ** $p < 0.01$, n.s., nonsignificant, as calculated by unpaired Student's *t* test. See also Figures S5 and S6.

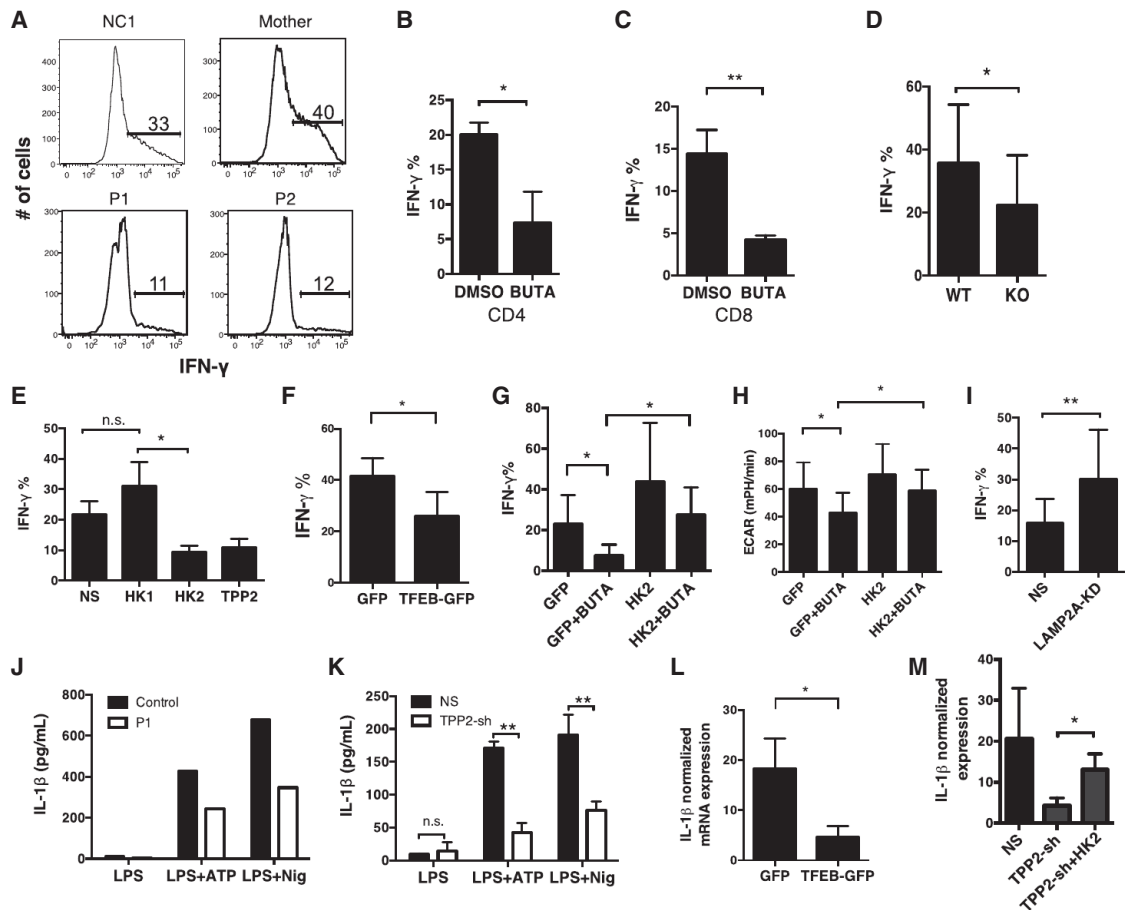
483–488) or the C-terminal QRFEK motif (residues 760–764) or an N-terminal fragment lacking both motifs. Coimmunoprecipitations showed that full-length HK2 interacted with HSC70, as did fragments containing either of two lysosomal targeting motifs (F2, F3); however, coimmunoprecipitation of the N-terminal fragment lacking both motifs (F1) was decreased (Figure 5F). Moreover, given that overexpression of the HK2 fragment F2 showed partial degradation, its interaction with HSC70 was

probably underestimated. Deletion of both lysosomal-targeting motifs resulted in a nondegradable HK2 protein that was expressed to higher levels (Figure 5G). We also considered the possibility that macroautophagy contributed to the loss of HK2 in TPPII-deficient cells but observed only slightly increased expression of LC3-II, a marker of autophagosomes, albeit levels increased with starvation (Figures S5I and S3G). Furthermore, normal cells, under the stress of amino acid



Data in (E) are represented as mean \pm SD from three independent experiments. See also Figure S5.

Oxidative phosphorylation is required for early activation of naive T cells, and T cell proliferation can be fueled by either oxidative phosphorylation or aerobic glycolysis (Chang et al., 2013). Accordingly, the preservation of oxidative phosphorylation in TPPII-deficient T cells (Figures 4G and 4H) was consistent with



their normal upregulation of activation markers and near normal proliferation after TCR stimulation (Figures S6A–S6G). However, a rapid transition to aerobic glycolysis is required to meet the increased biosynthetic and bioenergetic needs of effector functions in activated T cells, especially the acquisition of IFN- γ and IL-2 production in CD4 T cells (Chang et al., 2013). We therefore predicted that the glycolytic defect in TPP2-deficient T cells (Figures 4A–4D and 4F) would reduce their ability to produce effector

cytokines. As expected, we found that naive CD4 cells from P1 and P2 expressed less IFN- γ after activation (Figure 6A). Inhibiting TPP2 by BUTA in activated CD4 or CD8 T cells from healthy normal control subjects reduced IFN- γ posttranscriptionally (Figures 6B, 6C, and S7A) and also reduced IL-2 (Figures S7B and S7C), with no effects on TNF- α production (Figure S7D). Moreover, mouse *Tpp2* KO naive CD4 T cells expressed less IFN- γ even when differentiated under T_H1 polarizing conditions

with exogenously added IL-2 (Figure 6D). These effector functions were specifically decreased because other cytokines such as IL-4 and IL-17 were unaffected by TPPII deficiency (Figures S7E and S7F). Consistent with our metabolic model, TPPII inhibition of IFN- γ production was recapitulated by siRNA silencing of *HK2*, but not *HK1* (Figure 6E), or by overexpressing TFEB, which decreases HK2 protein (Figures 6F and S5G). Chloroquine did not restore glycolysis or cytokine production, probably because undegraded HK2 remained sequestered and nonfunctional within lysosomes (Figure S5F and data not shown). Importantly, overexpression of WT HK2 in TPPII-inhibited T cells restored both glycolysis and IFN- γ expression (Figures 6G, 6H, and S7G). Rescue also occurred when a nondegradable mutant HK2 (lacking both lysosomal targeting motifs) was overexpressed, despite its partially impaired enzymatic activity (Figures S7H and S7I). Finally, when the receptor for CMA substrates was reduced using *LAMP2A* siRNA, HK2 levels and IFN- γ production were increased in TPPII-replete cells (Figures 6I and S5G), suggesting that CMA also tonically regulates HK2. Thus, reduced glycolysis in TPPII deficiency results from increased lysosomal degradation of HK2 via CMA and leads to impaired T cell function.

Aerobic glycolysis contributes not only to adaptive immunity, but also to innate immunity (Tannahill et al., 2013). Thus, we examined macrophage inflammatory responses after priming with the TLR agonist LPS and activating with ATP or nigericin. Monocyte-derived macrophages from P1 secreted less IL-1 β than control (Figure 6J), as did *TPP2* shRNA-transduced THP1 cells (Figure 6K). This inhibition was likely transcriptional, as shown by reduced *IL1B* mRNA levels upon LPS priming of TPPII-deficient cells (Figures S7J and S7K) and could be recapitulated by overexpressing TFEB (Figures 6L and S5G). By contrast, *TNFA* mRNA levels (TNF- α) were less affected (Figures S7J and S7K). These defects would likely contribute to impaired responses to microbes or sterile tissue damage. HK2 overexpression in the *TPP2* shRNA-transduced THP1 cells restored their IL-1 β expression (Figures 6M, S7L, and S7M), similar to its rescue of IFN- γ production in T cells. Together, these results define autosomal recessive TPPII deficiency as a metabolic cause of primary immune deficiency in which reduced aerobic glycolysis leads to defects in both adaptive and innate immunity.

DISCUSSION

Despite its role as a peptidase downstream of the proteasome, previous studies of TPPII have focused on its ability to process antigenic peptides for MHC class I presentation. We now show that this giant cytoplasmic proteolytic complex has a more profound influence on overall amino acid homeostasis. Our data indicate that TPPII participates in recycling cellular proteins into amino acids and that extracellular amino acid transport inefficiently compensates for its loss unless supraphysiological levels are provided. When TPPII function is inhibited, free intracellular amino acids are transiently decreased and lysosomal activity is elevated via mTOR and TFEB as a compensatory mechanism to restore amino acid homeostasis. However, the altered amino acid equilibrium in TPPII-deficient cells remains fragile to extracellular nutrient shortage, which likely occurs

intermittently in vivo despite the lysosomal compensation at baseline. Consistent with previous reports that proteasomal inhibition increases lysosome abundance (Rideout et al., 2004; Ryh nen et al., 2009), our work now reveals the essential balance between the proteasomal and lysosomal degradation pathways for normal cell performance and that TPPII is an important linchpin in this process.

TPPII-deficient cells with intact lysosomal function maintain amino acid homeostasis at the cost of reduced metabolic fitness. We observed that lysosomal hyperproliferation causes HK2 degradation through HSC70-chaperoned lysosomal degradation. The HK2 isoenzyme is upregulated in both lymphocytes and tumor cells that utilize aerobic glycolysis (the Warburg effect) and differs from HK1, which is constitutively expressed (Robey and Hay, 2006). Because hexokinases perform a rate-limiting step in glycolysis, the targeted degradation of HK2 strongly impedes glycolytic flux in TPPII-deficient cells. Several glycolytic enzymes are degraded through CMA, which increases during amino acid starvation (Dice, 1990; Kaushik et al., 2011). CMA of HK2 was not recognized previously, possibly because an experimental model in which HK2 is not induced was used (Kon et al., 2011). Glycolysis can also be regulated by PFKP and PKM2, so the latter's decrease upon acute TPPII inhibition may further compromise glycolysis. Nevertheless, we show that, when TPPII is chronically lacking, HK2 consumption alone can explain the molecular regulation of glycolysis. It is likely that the pathologies caused by TPPII deficiency in cells normally dependent on HK2 are related to the cell-specific functions that are sensitive to glycolysis.

TPPII is widely expressed in different tissues but at relatively higher levels in the immune system and in tumors such as Burkitt's lymphoma (B low et al., 1986; Gavioli et al., 2001). These cell types sustain high levels of aerobic glycolysis. Moreover, HK2 but not HK1, promotes growth of tumor cells through effects on aerobic glycolysis and downstream biosynthetic pathways (Gershon et al., 2013; Patra et al., 2013; Wolf et al., 2011a). In the immune system, aerobic glycolysis is required for peripheral T cell responses and for thymocyte development (Greiner et al., 1994). Activated macrophages and dendritic cells also require aerobic glycolysis to generate biosynthetic intermediates, including fatty acids necessary for production of membranes and proinflammatory cytokines (Everts et al., 2014; Tannahill et al., 2013). In TPPII deficiency, adaptive and innate immune cells function poorly in elaborating IFN- γ , IL-2, and IL-1 β , which explains the patients' susceptibility to a broad range of pathogens. Furthermore, the severe immune dysregulation characteristic of this disorder, including autoimmunity, may reflect the differential sensitivity of effector functions to impaired aerobic glycolysis and secondary metabolic defects.

The role of metabolic processes in immune cell functions has been mainly elucidated through in vitro studies using pharmacological agents such as 2-DG or mice as tools in which transcriptional regulators such as HIF1 α were selectively ablated. By studying a new human immunodeficiency, we now better understand the requirements and physiological consequences of metabolic reprogramming on immune functions in vivo. We find that TPPII deficiency produced less impairment of glycolytic flux than 2-DG treatment (Lampidis et al., 2006), consonant with

its selective effect on HK2 that spares oxidative phosphorylation and, hence, T cell proliferation (Cham et al., 2008; Miller et al., 1994). Thus, our data showing unchanged IL-4 and IL-17 can be reconciled with previous reports that glycolytic blockade by 2-DG in mice interrupts not only IFN- γ , but also IL-4 and IL-17 production (Shi et al., 2011).

Finally, although hematopoietic stem cell transplantation would be expected to cure the immune abnormalities in patients with TPPII deficiency, neurometabolic correction presents a greater challenge. In the brain, TPPII degrades neuropeptides to regulate satiety (Rose et al., 1996; Wilson et al., 1993). However, HK2 is also dynamically expressed in the developing brain, mirroring the pattern of whole-brain glucose consumption, which is high during early childhood before decreasing by adulthood (Gershon et al., 2013; Wolf et al., 2011b). Increased utilization of aerobic glycolysis, which predominates in brain regions associated with transcriptional neoteny, may contribute to synapse formation and growth (Goyal et al., 2014). Thus, our finding that TPPII inhibition impaired glycolysis in a neuroblastoma cell line suggests the intriguing hypothesis that the developmental delay in TPPII deficiency is mechanistically linked to defective aerobic glycolysis. Because of this unusual constellation of features, we propose that this disorder be called “TPPII-related immunodeficiency, autoimmunity, and neurodevelopmental delay with impaired glycolysis and lysosomal expansion” (TRI-ANGLE) disease.

EXPERIMENTAL PROCEDURES

Study Subjects and Mice

Patients and their relatives provided written informed consent to participate in research protocols approved by the NIAID, NIH Institutional Review Board (NCT00246857) or the Newcastle and North Tyneside 1 Research Ethics Committee. Whole-blood samples, primary dermal fibroblast cultures, and buffy coat cells were obtained. Mice were bred, purchased, and used under animal study protocols approved by the NIAID Animal Care Use Committee. TppII-deficient mice were from Kenneth Rock (University of Massachusetts, Worcester) (Kawahara et al., 2009). Details are provided in the Extended Experimental Procedures.

Genomic Analyses

WES of genomic DNA was performed for P1 and P3 and genome-wide linkage scans for P3 and P4. Details of analyses, including alignment, variant calling, and filtering, are provided in the Extended Experimental Procedures. Mutations were confirmed by Sanger sequencing. WES data are deposited in dbGaP.

Molecular Modeling

The HsTPPII spindle complex was used to model the G500D mutation. Details are provided in the Extended Experimental Procedures.

Cell Isolation and Culture

Human PBMC, human peripheral blood T cells (including naive CD4 or CD8 T cells), human monocyte-derived macrophages, mouse splenic naive CD4 or CD8 T cells, MEF, human fibroblasts, THP1, 293T, A549, SH-SY5Y, Jurkat, and HeLa cell lines were used. Details are provided in the Extended Experimental Procedures.

Immunoblotting

Immunoblotting was performed according to standard methods. To assess protein degradation rates, A549 cells were transfected with myc-tagged mammalian expression constructs; 24 hr later, cycloheximide (10 μ g/ml) was

added and cells cultured up to 20 more hr before lysis and immunoblotting for myc. Details are provided in the Extended Experimental Procedures.

TPPII Enzymatic Activity

Cell lysates were incubated with Ala-Ala-Phe-4-methylcoumaryl-7-amide (AAF-AMC) (Enzo Life Sciences) and fluorescence emission measured. Details are provided in the Extended Experimental Procedures.

Native In-Gel AAF-AMC Cleavage and Immunoblotting

Cell lysates were separated by NativePAGE, followed by either transfer to PVDF membranes for immunoblotting or in-gel incubation with AAF-AMC to assess TPPII enzymatic activity. Details are provided in the Extended Experimental Procedures.

Cell Treatments

Human fibroblasts were treated in medium containing reduced amino acids and MEF in amino acid free media for 6 hr (USBiological Life Sciences). PBMC were stimulated with anti-human CD2/3/28 beads (Miltenyi Biotec) for T cell activation and CFSE-labeled cell proliferation assays. Cells were treated with the TPPII-specific inhibitor butabindide (Santa Cruz Biotechnology; Tocris Bioscience) at 200 or 250 μ M. Purified human or mouse naive CD4 T or CD8 cells were stimulated with anti-CD3/CD28 antibodies; subset-polarizing recombinant cytokines and antibodies against mouse cytokines were added for T helper cell differentiation studies. Human monocyte-derived macrophages were stimulated with LPS (100 ng/ml) for 6 hr and then ATP (5 mM) or nigericin (10 mM) for 30 more min (all from Invivogen). THP1 cells were stimulated with phorbol 12-myristate 13-acetate (PMA, 100 ng/ml; Sigma-Aldrich) for 3 hr and then LPS (1 μ g/ml) for 3–48 hr and, in some cases, ATP (5 mM) or nigericin (10 mM) for 30 more min. Cells were treated for 6 hr with MG132 (5 mM, Enzo Life Science), chloroquine (100 mM), or ammonium chloride (20 mM) plus leupeptin (100 μ M) (both from Sigma-Aldrich). Details are provided in the Extended Experimental Procedures.

Intracellular Free Amino Acid Analyses

Samples were analyzed for intracellular free amino acids by Waters ACQUITY ultraperformance liquid chromatography (UPLC) after AccQ-Tag derivatization or by optical absorbance after reacting with ninhydrin. Details are provided in the Extended Experimental Procedures.

Lentiviral Transductions and Transient Transfections

THP1 cells were stably transduced with lentivirus encoding shRNA against human *TPP2* or control shRNA. Human CD4 naive T cells were Amara nucleofected with 200 nM siRNA against *LAMP2A*, *HK1*, or *HK2* (Life Technologies). Mammalian expression constructs encoding tagged forms of TFEB, HK1, HK2, PKM2, or GAPDH were transfected by nucleofection or Turbofect. Details are provided in the Extended Experimental Procedures.

Transmission Electron Microscopy

Transmission electron microscopy of human cycling T cells was performed after fixation in 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer (pH 7.4). Details are provided in the Extended Experimental Procedures.

Lysosomal Imaging and Quantitation

The lysosome-specific dyes LysoTracker Red (Life Technology) and Lyso-ID (Enzo Life Sciences) were used to stain intracellular lysosomes for confocal microscopy. Staining for T cell surface markers, LysoID, and LC3 for autophagosomes were performed for ImageStream analyses. Details are provided in the Extended Experimental Procedures.

Lysosomal Enzymatic Activity

Preparations enriched in lysosomes were collected from mouse liver tissue homogenates using the Lysosome Isolation Kit with density gradient ultracentrifugation. The fraction containing the majority of the lysosomes, as determined by immunoblotting of lysosomal proteins, was used to measure lysosomal acid phosphatase and β -N-acetyl-glucosaminidase activities (all from Sigma-Aldrich). Details are provided in the Extended Experimental Procedures.

Quantitative RT-PCR

Total RNA was isolated with the RNeasy kit (QIAGEN). cDNA was synthesized from 1 μ g of RNA using the iScript cDNA Synthesis Kit (Bio-Rad). All quantitative RT-PCR was performed by the SYBR green method on 7900HT machine (ABI), using 200 pg of cDNA per reaction. The expression of mRNA for genes of interest was normalized to that of β -actin or HPRT. Primer sequences are provided in the Extended Experimental Procedures.

TFEB Subcellular Localization

At 24 hr after transfecting a C-terminal EGFP-fused TFEB plasmid (Roczniak-Ferguson et al., 2012), cells were fixed with 4% paraformaldehyde in PBS, their nuclei stained with DAPI (Enzo Life Sciences), and imaged by confocal microscopy, as described above ("Lysosomal Imaging and Quantitation"). For analysis of percentage of nuclear colocalization, single cells were analyzed by ImaRis with threshold settings of 100 in both GFP and DAPI channels. Ten TFEB-GFP-positive cells were analyzed per experimental group. Details are provided in the Extended Experimental Procedures.

Metabolic Analyses

Glycolysis was evaluated by measuring ECAR at baseline and after adding glucose (10 mM), oligomycin (10 μ M), and 2-deoxy-D-glucose (100 mM). Oxidative phosphorylation was evaluated by measuring OCR at baseline and after adding oligomycin (5 μ M), carbonyl cyanide p-trifluoromethoxyphenylhydrazone (FCCP, 1.5 μ M), and rotenone (100 nM) plus antimycin (1 μ M). Measurements were made on an XF-96 Extracellular Flux Analyzer using the XF Glycolysis or Cell Mito stress test kits (Seahorse Bioscience). Glycolytic flux was also measured by monitoring conversion of [5- 3 H]-glucose to [5- 3 H]-water. Details are provided in the Extended Experimental Procedures.

Coimmunoprecipitations

At 24 hr after transfection of SH-SY5Y or 293T cells with GFP-HSC70 and myc-HK2 expression plasmids, supernatants from cell lysates were incubated for 8 hr at 4°C with Dynal beads (Life Technology) plus anti-myc tag antibody (Cell Signaling Technology). Beads were washed with ice-cold lysis buffer and the bound proteins eluted for SDS-PAGE and immunoblotting, as described above ("Immunoblotting"). Details are provided in the Extended Experimental Procedures.

Flow Cytometric Analyses

Standard flow cytometry methods were used to evaluate surface markers of T cell activation (CD25, CD69), intracellular cytokines (IFN- γ , IL-2, IL-4, IL-17, TNF- α), mTOR activity (phospho-S6), and cell division (CFSE dilutions). Data were acquired on a BD LSRII instrument using FACSDiva software and were analyzed using FlowJo software (Tree Star). Details are provided in the Extended Experimental Procedures.

ELISA Quantitation of Cytokines

IL-1 β was measured in conditioned media after treating monocyte-derived macrophages or THP1 cells as described above, using an IL-1 β ELISA kit (Biolegend). Details are provided in the Extended Experimental Procedures.

HK2 Rescue

At 24 hr after transfection with a carboxyl-tagged GFP fusion protein of HK2 or derivatives thereof, human naive CD4 T cells were stimulated for 4 days with anti-CD3/CD28 antibodies. HK2-transfected THP1 cells were treated with PMA (100 ng/ml) for 3 hr and then LPS (1 μ g/ml) for 48 more hr. Details are provided in the Extended Experimental Procedures.

Statistical Analyses

The Prism 5 software package (GraphPad) was used to calculate p values using the unpaired two-tailed Student's t test, one-sample t test, one-way ANOVA, or two-way ANOVA.

ACCESSION NUMBERS

The human exome sequencing data has been deposited to dbGaP with the accession number phs000848.v1.p1.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and one data file and can be found with this article at <http://dx.doi.org/10.1016/j.cell.2014.12.001>.

AUTHOR CONTRIBUTIONS

W.L. generated *TPP2* shRNA THP1 cells, measured intracellular free amino acids and HK activity, performed HK2 rescue and coimmunoprecipitations, and quantified cytokine production by flow cytometry and ELISA. Y.Z. and, independently, D.O.M., L.N.R., M.F.S.K., N.V.M., and H.G. discovered the patients' *TPP2* mutations and evaluated their effects on expression. D.O.M. performed native gel electrophoresis and evaluated TPPII enzymatic activity. D.O.M. discovered and W.L. characterized defective glycolysis. W.L., D.O.M., H.J., and Q.Z. evaluated T cell activation and proliferation. H.J. discovered and W.L., S.S., A.K.S., V.I.K., B.C., and Q.Z. characterized the lysosomal expansion. W.L., H.J., and H.M.M. performed immunoblotting of glycolytic enzymes. L.Z. performed radiotracer studies of glycolysis. W.L. and Q.Z. performed confocal microscopy. Y.Z., W.L., H.J., and H.M.M. performed qRT-PCR experiments. H.J. generated HK2 mutant expression constructs. B.C., N.R., V.I.K., and W.L. assessed mTOR activity and autophagy. J.H.L. modeled the effects of the mutation on protein structure. S.E.T., T.P., S.J.H., J.M.H., S.H., and A.J.C. provided clinical care and collected clinical data to characterize the patients. H.F.M. coordinated clinical study protocol and sample collection. H.C.S., M.J.L., and S.H. planned and supervised the experimental work and data analyses. Y.Z., W.L., H.C.S., and S.H. prepared the manuscript. All authors discussed and commented on the manuscript.

ACKNOWLEDGMENTS

We thank the following people: for technical help, David Dorward, Louise Tee, Sonia Majri, Matt Biancalana, Alex Leney-Greene, Sam Drain, and Yaobo Xu; for clinical support, Angela Wang, Emily Jewel, Shanaz Pasha, and colleagues at the BC Children's Hospital and the Great North Children's Hospital; for sharing reagents, data, and equipment, Kenneth Rock, Beate Rockel, Thomas von Zglinicki, and Richard Siegel; for help with genetic investigations, Eamonn Maher and the Newcastle Exome Consortium, especially John Loughlin and Patrick Chinnery; for advice and critically reading the manuscript, Karin Engelhardt, Richard Siegel, Fengyi Wan, and Li Yu; and the patients and their families for participating in this study. This work was supported by the Intramural Research Program of the NIAID, NIH (H.C.S., M.J.L.); the UK MRC (S.H.); the Sir Jules Thorn Charitable Trust (S.H., M.F.S.K.); the NIHR Newcastle Biomedical Research Centre and the Newcastle upon Tyne Hospitals NHS Charity (S.H., M.F.S.K., V.I.K.); UK BBSRC (B.C., V.I.K.); NIHR Oxford Biomedical Research Centre (S.S., A.K.S.); and the Canadian Institutes of Health Research (S.E.T., MOP-133691). S.E.T. is a Clinical Scholar of the Michael Smith Foundation for Health Research.

Received: September 19, 2014

Revised: November 17, 2014

Accepted: November 30, 2014

Published: December 18, 2014

REFERENCES

- Bålöw, R.M., Tomkinson, B., Ragnarsson, U., and Zetterqvist, O. (1986). Purification, substrate specificity, and classification of tripeptidyl peptidase II. *J. Biol. Chem.* 261, 2409–2417.
- Bar-Peled, L., and Sabatini, D.M. (2014). Regulation of mTORC1 by amino acids. *Trends Cell Biol.* 24, 400–406.
- Barnes, D.M., Calvert, C.C., and Klasing, K.C. (1992). Source of amino acids for tRNA acylation. Implications for measurement of protein synthesis. *Biochem. J.* 283, 583–589.
- Cham, C.M., Driessens, G., O'Keefe, J.P., and Gajewski, T.F. (2008). Glucose deprivation inhibits multiple key gene expression events and effector functions in CD8+ T cells. *Eur. J. Immunol.* 38, 2438–2450.

- Chang, C.H., Curtis, J.D., Maggi, L.B., Jr., Faubert, B., Villarino, A.V., O'Sullivan, D., Huang, S.C., van der Windt, G.J., Blagih, J., Qiu, J., et al. (2013). Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell* 153, 1239–1251.
- Ciechanover, A. (2005). Intracellular protein degradation: from a vague idea thru the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting. *Cell Death Differ.* 12, 1178–1190.
- Dice, J.F. (1990). Peptide sequences that target cytosolic proteins for lysosomal proteolysis. *Trends Biochem. Sci.* 15, 305–309.
- Everts, B., Amiel, E., Huang, S.C., Smith, A.M., Chang, C.H., Lam, W.Y., Redmann, V., Freitas, T.C., Blagih, J., van der Windt, G.J., et al. (2014). TLR-driven early glycolytic reprogramming via the kinases TBK1-IKK ϵ supports the anabolic demands of dendritic cell activation. *Nat. Immunol.* 15, 323–332.
- Gavioli, R., Frisan, T., Vertuani, S., Bornkamm, G.W., and Masucci, M.G. (2001). c-myc overexpression activates alternative pathways for intracellular proteolysis in lymphoma cells. *Nat. Cell Biol.* 3, 283–288.
- Gershon, T.R., Crowther, A.J., Tikunov, A., Garcia, I., Annis, R., Yuan, H., Miller, C.R., Macdonald, J., Olson, J., and Deshmukh, M. (2013). Hexokinase-2-mediated aerobic glycolysis is integral to cerebellar neurogenesis and pathogenesis of medulloblastoma. *Cancer Metab.* 1, 2.
- Goyal, M.S., Hawrylycz, M., Miller, J.A., Snyder, A.Z., and Raichle, M.E. (2014). Aerobic glycolysis in the human brain is associated with development and neonatal gene expression. *Cell Metab.* 19, 49–57.
- Greiner, E.F., Guppy, M., and Brand, K. (1994). Glucose is essential for proliferation and the glycolytic enzyme induction that provokes a transition to glycolytic energy production. *J. Biol. Chem.* 269, 31484–31490.
- Huai, J., Firat, E., Nil, A., Million, D., Gaedicke, S., Kanzler, B., Freudenberg, M., van Endert, P., Kohler, G., Pahl, H.L., et al. (2008). Activation of cellular death programs associated with immunosenescence-like phenotype in TPPII knockout mice. *Proc. Natl. Acad. Sci. USA* 105, 5177–5182.
- Kaushik, S., Bandyopadhyay, U., Sridhar, S., Kiffin, R., Martinez-Vicente, M., Kon, M., Orenstein, S.J., Wong, E., and Cuervo, A.M. (2011). Chaperone-mediated autophagy at a glance. *J. Cell Sci.* 124, 495–499.
- Kawahara, M., York, I.A., Hearn, A., Farfan, D., and Rock, K.L. (2009). Analysis of the role of tripeptidyl peptidase II in MHC class I antigen presentation in vivo. *J. Immunol.* 183, 6069–6077.
- Kon, M., Kiffin, R., Koga, H., Chapochnick, J., Macian, F., Varticovski, L., and Cuervo, A.M. (2011). Chaperone-mediated autophagy is required for tumor growth. *Sci. Transl. Med.* 3, 109ra117.
- Korolchuk, V.I., Menzies, F.M., and Rubinsztein, D.C. (2010). Mechanisms of cross-talk between the ubiquitin-proteasome and autophagy-lysosome systems. *FEBS Lett.* 584, 1393–1398.
- Lampidis, T.J., Kurtoglu, M., Maher, J.C., Liu, H.P., Krishan, A., Sheft, V., Szymanski, S., Fokt, I., Rudnicki, W.R., Ginalska, K., et al. (2006). Efficacy of 2-halogen substituted D-glucose analogs in blocking glycolysis and killing “hypoxic tumor cells”. *Cancer Chemother. Pharmacol.* 58, 725–734.
- MacIver, N.J., Michalek, R.D., and Rathmell, J.C. (2013). Metabolic regulation of T lymphocytes. *Annu. Rev. Immunol.* 31, 259–283.
- McKay, R.M., McKay, J.P., Suh, J.M., Avery, L., and Graff, J.M. (2007). Tripeptidyl peptidase II promotes fat formation in a conserved fashion. *EMBO Rep.* 8, 1183–1189.
- Miller, E.S., Klinger, J.C., Akin, C., Koebel, D.A., and Sonnenfeld, G. (1994). Inhibition of murine splenic T lymphocyte proliferation by 2-deoxy-D-glucose-induced metabolic stress. *J. Neuroimmunol.* 52, 165–173.
- Patra, K.C., Wang, Q., Bhaskar, P.T., Miller, L., Wang, Z., Wheaton, W., Chandel, N., Laakso, M., Muller, W.J., Allen, E.L., et al. (2013). Hexokinase 2 is required for tumor initiation and maintenance and its systemic deletion is therapeutic in mouse models of cancer. *Cancer Cell* 24, 213–228.
- Pearce, E.L., and Pearce, E.J. (2013). Metabolic pathways in immune cell activation and quiescence. *Immunity* 38, 633–643.
- Reits, E., Neijssen, J., Herberths, C., Benchuijsen, W., Janssen, L., Drijfhout, J.W., and Neeffes, J. (2004). A major role for TPPII in trimming proteasomal degradation products for MHC class I antigen presentation. *Immunity* 20, 495–506.
- Rideout, H.J., Lang-Rollin, I., and Stefanis, L. (2004). Involvement of macroautophagy in the dissolution of neuronal inclusions. *Int. J. Biochem. Cell Biol.* 36, 2551–2562.
- Robey, R.B., and Hay, N. (2006). Mitochondrial hexokinases, novel mediators of the antiapoptotic effects of growth factors and Akt. *Oncogene* 25, 4683–4696.
- Roczniak-Ferguson, A., Petit, C.S., Froehlich, F., Qian, S., Ky, J., Angarola, B., Walther, T.C., and Ferguson, S.M. (2012). The transcription factor TFEB links mTORC1 signaling to transcriptional control of lysosome homeostasis. *Sci. Signal.* 5, ra42.
- Rose, C., Vargas, F., Facchinetti, P., Bourgeat, P., Bambal, R.B., Bishop, P.B., Chan, S.M., Moore, A.N., Ganellin, C.R., and Schwartz, J.C. (1996). Characterization and inhibition of a cholecystokinin-inactivating serine peptidase. *Nature* 380, 403–409.
- Ryhänen, T., Hyttinen, J.M., Kopitz, J., Rilla, K., Kuusisto, E., Mannermaa, E., Viiri, J., Holmberg, C.I., Immonen, I., Meri, S., et al. (2009). Crosstalk between Hsp70 molecular chaperone, lysosomes and proteasomes in autophagy-mediated proteolysis in human retinal pigment epithelial cells. *J. Cell. Mol. Med.* 13 (9B), 3616–3631.
- Schönege, A.M., Villa, E., Förster, F., Hegerl, R., Peters, J., Baumeister, W., and Rockel, B. (2012). The structure of human tripeptidyl peptidase II as determined by a hybrid approach. *Structure* 20, 593–603.
- Schutz, Y. (2011). Protein turnover, ureagenesis and gluconeogenesis. *Int. J. Vit. Nutr. Res.* 81, 101–107.
- Settembre, C., Di Malta, C., Polito, V.A., Garcia-Arencibia, M., Vetrini, F., Erdin, S., Erdin, S.U., Huynh, T., Medina, D., Colella, P., et al. (2011). TFEB links autophagy to lysosomal biogenesis. *Science* 332, 1429–1433.
- Shi, L.Z., Wang, R., Huang, G., Vogel, P., Neale, G., Green, D.R., and Chi, H. (2011). HIF1 α -dependent glycolytic pathway orchestrates a metabolic checkpoint for the differentiation of TH17 and Treg cells. *J. Exp. Med.* 208, 1367–1376.
- Stentz, F.B., and Kitabchi, A.E. (2004). Transcriptome and proteome expression in activated human CD4 and CD8 T-lymphocytes. *Biochem. Biophys. Res. Commun.* 324, 692–696.
- Tannahill, G.M., Curtis, A.M., Adamik, J., Palsson-McDermott, E.M., McGettrick, A.F., Goel, G., Frezza, C., Bernard, N.J., Kelly, B., Foley, N.H., et al. (2013). Succinate is an inflammatory signal that induces IL-1 β through HIF-1 α . *Nature* 496, 238–242.
- Taylor, P.M. (2014). Role of amino acid transporters in amino acid sensing. *Am. J. Clin. Nutr.* 99, 223S–230S.
- Tomkinson, B. (1999). Tripeptidyl peptidases: enzymes that count. *Trends Biochem. Sci.* 24, 355–359.
- Valdor, R., Mocholi, E., Botbol, Y., Guerrero-Ros, I., Chandra, D., Koga, H., Gravekamp, C., Cuervo, A.M., and Macian, F. (2014). Chaperone-mediated autophagy regulates T cell responses through targeted degradation of negative regulators of T cell activation. *Nat. Immunol.* 15, 1046–1054.
- Wang, R., Dillon, C.P., Shi, L.Z., Milasta, S., Carter, R., Finkelstein, D., McCormick, L.L., Fitzgerald, P., Chi, H., Munger, J., and Green, D.R. (2011). The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity* 35, 871–882.
- Wilson, C., Gibson, A.M., and McDermott, J.R. (1993). Purification and characterization of tripeptidylpeptidase-II from post-mortem human brain. *Neurochem. Res.* 18, 743–749.
- Wolf, A., Agnihotri, S., Micallef, J., Mukherjee, J., Sabha, N., Cairns, R., Hawkins, C., and Guha, A. (2011a). Hexokinase 2 is a key mediator of aerobic glycolysis and promotes tumor growth in human glioblastoma multiforme. *J. Exp. Med.* 208, 313–326.
- Wolf, A., Agnihotri, S., Munoz, D., and Guha, A. (2011b). Developmental profile and regulation of the glycolytic enzyme hexokinase 2 in normal brain and glioblastoma multiforme. *Neurobiol. Dis.* 44, 84–91.

Acetate Dependence of Tumors

Sarah A. Comerford,^{1,7} Zhiguang Huang,^{2,7} Xinlin Du,² Yun Wang,² Ling Cai,² Agnes K. Witkiewicz,³ Holly Walters,¹ Mohammed N. Tantawy,^{4,5} Allie Fu,⁴ H. Charles Manning,^{4,5,6} Jay D. Horton,¹ Robert E. Hammer,² Steven L. McKnight,^{2,8,*} and Benjamin P. Tu^{2,8,*}

¹Department of Molecular Genetics, UT Southwestern Medical Center, Dallas, TX 75390, USA

²Department of Biochemistry, UT Southwestern Medical Center, Dallas, TX 75390, USA

³Department of Pathology, UT Southwestern Medical Center, Dallas, TX 75390, USA

⁴Vanderbilt University Institute of Imaging Science, Vanderbilt University Medical Center, Nashville, TN 37232, USA

⁵Department of Radiology and Radiological Sciences, Vanderbilt University Medical Center, Nashville, TN 37232, USA

⁶Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN 37232, USA

⁷Co-first author

⁸Co-senior author

*Correspondence: steven.mcknight@utsouthwestern.edu (S.L.M.), benjamin.tu@utsouthwestern.edu (B.P.T.)

<http://dx.doi.org/10.1016/j.cell.2014.11.020>

SUMMARY

Acetyl-CoA represents a central node of carbon metabolism that plays a key role in bioenergetics, cell proliferation, and the regulation of gene expression. Highly glycolytic or hypoxic tumors must produce sufficient quantities of this metabolite to support cell growth and survival under nutrient-limiting conditions. Here, we show that the nucleocytosolic acetyl-CoA synthetase enzyme, ACSS2, supplies a key source of acetyl-CoA for tumors by capturing acetate as a carbon source. Despite exhibiting no gross deficits in growth or development, adult mice lacking ACSS2 exhibit a significant reduction in tumor burden in two different models of hepatocellular carcinoma. ACSS2 is expressed in a large proportion of human tumors, and its activity is responsible for the majority of cellular acetate uptake into both lipids and histones. These observations may qualify ACSS2 as a targetable metabolic vulnerability of a wide spectrum of tumors.

INTRODUCTION

Cell growth and proliferation are intimately coordinated with metabolism. Potentially distinct differences in metabolism between normal and cancerous cells have sparked a renewed interest in targeting metabolic enzymes as an approach to the discovery of new anticancer therapeutics. The metabolic strategies utilized by cancer cells to enhance proliferative capacity under nutrient-limiting conditions remain controversial and poorly understood. It has thus been unclear as to which aspects of cell metabolism might represent a realistic, targetable vulnerability of tumors relative to normal cells and tissues.

We recently found that prototrophic yeast cells monitor intracellular levels of acetyl-CoA in order to commit to a new round of cell division (Cai et al., 2011; Shi and Tu, 2013). Acetyl-CoA is a key intermediate of carbon sources, which not only fuels ATP production via the TCA cycle, but also functions

as an essential building block for the synthesis of fatty acids and sterols. When yeast cells commit to cell division, they significantly enhance the production of acetyl-CoA. Elevated levels of acetyl-CoA induce acetylation of histones on a set of more than 1,000 genes critical for cell growth (Cai et al., 2011). This battery of “growth genes” includes virtually all genes important for ribosome biogenesis, protein translation, and amino acid biosynthesis. Transcription of the key G1 cyclin (*CLN3*) that gates entry of yeast cells into the cell division cycle is also dependent upon the ability of cells to substantially enhance the intracellular abundance of acetyl-CoA (Shi and Tu, 2013). Thus, in budding yeast, acetyl-CoA is a sentinel metabolite that regulates transcription of growth genes via epigenetic modification of chromatin (Cai and Tu, 2011; Kaelin and McKnight, 2013).

The strict dependence of yeast cells on acetyl-CoA for cell growth and proliferation prompted us to examine whether acetyl-CoA might also be rate limiting for mammalian cell growth. In well-fed mammalian cells, the acetyl-CoA used for lipid synthesis and histone acetylation is primarily supplied by mitochondrially derived citrate (Srere, 1959; Wellen et al., 2009). This metabolite is enzymatically converted into acetyl-CoA via ATP citrate lyase (ACLY) (Srere, 1959; Srere and Lipmann, 1953). Cells grown under the nutrient-unlimited conditions of tissue culture medium also make acetyl-CoA via citrate consumption. By contrast, the nutrient-limiting conditions of tumor growth in animals and humans bring into question what pathways might be primarily utilized for acetyl-CoA production. The phenomenon of aerobic glycolysis famously characterized by Otto Warburg described the truncation of glucose oxidation at pyruvate (Warburg, 1956a, 1956b). Instead of pyruvate being transported into mitochondria for conversion into acetyl-CoA by the pyruvate dehydrogenase complex, many cancer cells are highly glycolytic and preferentially convert pyruvate into lactate. If pyruvate fails to enter the TCA cycle in cancer cells, how is it that sufficient citrate is made for ACLY-mediated production of acetyl-CoA?

Several groups have recently demonstrated the conversion of glutamine into acetyl-CoA via the phenomenon of reductive carboxylation whereby the TCA cycle can be modified to run in reverse (Le et al., 2012; Leonardi et al., 2012; Metallo et al.,

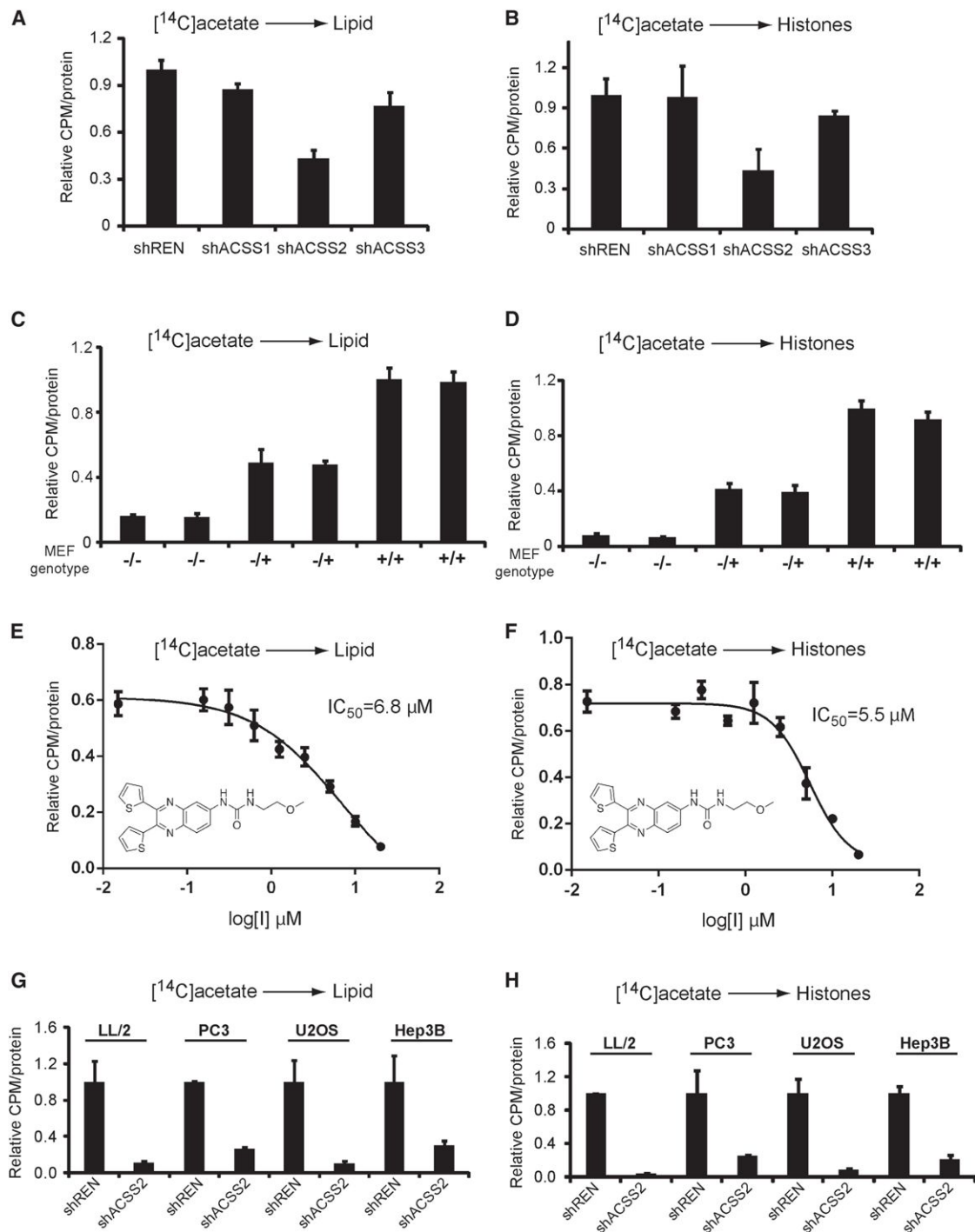


Figure 1. ACSS2 Is the Major Enzyme Required for Incorporation of Acetate into Lipids and Histones

(A) HepG2 cells with constitutive shRNA knockdown of ACSS1, ACSS2, ACSS3, or control (REN) were assayed for their ability to utilize [¹⁴C]acetate for lipid synthesis. Acetate must be converted to acetyl-CoA before it can be utilized as a metabolic substrate. Knockdown efficiency is shown in Figure S1 (mean ± SD, n = 3).

(B) HepG2 cells with constitutive shRNA knockdown of ACSS1, ACSS2, ACSS3, or control (REN) were assayed for their ability to utilize [¹⁴C]acetate for histone acetylation (mean ± SD, n = 3).

(C) Mouse embryonic fibroblasts (MEFs) were prepared from ACSS2 WT (+/+), heterozygous (-/+), and KO (-/-) mice and assayed for their ability to incorporate [¹⁴C]acetate into lipid. Note the -/- MEFs show very little [¹⁴C]acetate incorporation into lipid fractions. ACSS2 protein levels in the MEFs are shown in Figure S1 (mean ± SD, n = 3).

(legend continued on next page)

2012; Mullen et al., 2012; Wise et al., 2011). Whereas evidence supportive of reductive carboxylation has been obtained in studies of cancer cells grown in tissue culture, in vivo studies of primary human glioblastomas (GBMs) to date have revealed little or no catabolism of glutamine (Marin-Valencia et al., 2012). These GBMs instead exhibit substantive mitochondrial oxidation and a net synthesis of glutamine from glucose. Thus, the ability of glutamine to function as a source of acetyl-CoA in native tumors remains unclear.

These perplexing observations led us to consider alternative sources of acetyl-CoA for tumors in which, as a result of highly glycolytic or hypoxic metabolic environments, glucose-derived pyruvate is preferentially shunted toward lactate instead of acetyl-CoA. Budding yeast lack ATP citrate lyase and instead rely on a family of enzymes called acetyl-CoA synthetases (De Virgilio et al., 1992; Takahashi et al., 2006; van den Berg et al., 1996). Acetyl-CoA synthetases catalyze the synthesis of acetyl-CoA from acetate and CoA in an ATP-dependent reaction (Berg, 1956; Jones et al., 1953; Lipmann and Tuttle, 1945). We hypothesized that the mammalian versions of these acetyl-CoA synthetase enzymes might help cancer cells produce acetyl-CoA from acetate under the challenging growth conditions of solid tumors. Consistent with this idea, acetate could rescue histone acetylation in cell lines in which ACLY was knocked down, although the physiological relevance of acetate in mammalian cells was questioned (Wellen et al., 2009). However, a role for acetate in fueling tumor growth is supported by positron emission tomography (PET) imaging studies using [^{11}C]acetate wherein numerous clinical studies have documented avid acetate uptake in prostate, lung, liver, and brain cancers (Ho et al., 2003; Nomori et al., 2008; Oyama et al., 2002; Tsuchida et al., 2008). Indeed, in certain cases, [^{11}C]acetate PET imaging is more accurate and sensitive than [^{18}F]fluorodeoxyglucose (FDG) PET imaging, and some tumors are [^{11}C]acetate-positive yet FDG negative. These considerations have led to the proposal that acetyl-CoA synthetase enzymes could be important for [^{11}C]acetate uptake and tumor cell survival (Yoshii et al., 2009a, 2009b; Yun et al., 2009). In the accompanying article by Mashimo et al (2014) in this issue of *Cell*, acetate consumption by human tumors was recently confirmed by nuclear magnetic resonance (NMR) facilitated [^{13}C]acetate metabolic tracer experiments. Here, we report evidence that the nucleocytosolic ACSS2 enzyme is of critical importance for mammalian cells to utilize acetate as a source of acetyl-CoA, and that mice lacking this enzyme exhibit a substantial reduction in tumor burden in two genetic models of liver cancer.

RESULTS

ACSS2 Is Required for Acetate Uptake and Utilization in Mammalian Cells

The mammalian genome contains genes encoding three different enzymes capable of catalyzing the ATP-dependent synthesis of acetyl-CoA from acetate (Watkins et al., 2007). Two such enzymes, designated ACSS1 and ACSS3, are mitochondrial proteins (Fujino et al., 2001; Pérez-Chacón et al., 2009). The third, designated ACSS2, has been reported to be localized to both the cytoplasmic and nuclear compartments of mammalian cells (Ariannur et al., 2010; Luong et al., 2000). In order to assess the relative contributions of these enzymes for cellular utilization of acetate for either lipid synthesis or histone acetylation, RNAi agents were deployed to selectively silence their respective production. After having observed substantive, RNAi-mediated suppression of ACSS1, ACSS2, and ACSS3 (Figure S1 available online), cells were exposed to [^{14}C]acetate in order to measure incorporation of acetyl units into either lipids or histones. RNAi-mediated suppression of ACSS2 led to a more significant diminution in both lipid and histone assimilation of radiolabeled acetate than suppression of ACSS1 or ACSS3 (Figures 1A and 1B).

In order to pursue these observations in a more rigorous manner, mouse embryonic fibroblasts (MEFs) were prepared from embryos of mice bearing inactivating mutations in both alleles of ACSS2 (ACSS2 $^{-/-}$). Littermate embryos heterozygous for the ACSS2 mutation (ACSS2 $^{+/-}$), and wild-type littermates (ACSS2 $^{+/+}$), were also used to prepare MEFs (Figure S1). When exposed to [^{14}C]acetate, ACSS2 $^{-/-}$ MEFs showed a substantial deficit in label incorporation into both lipids and histones relative to ACSS2 $^{+/+}$ MEFs (Figures 1C and 1D). Cells heterozygous for ACSS2 (ACSS2 $^{+/-}$) revealed an intermediate decrement in [^{14}C]acetate uptake.

As a third means of testing the importance of ACSS2 for acetate uptake into mammalian cells, a high-throughput screen was conducted in search of chemical inhibitors of the enzyme. Purified, recombinant human ACSS2 enzyme was screened against roughly 200,000 drug-like chemicals housed within the UTSW compound file. The details of this screen are presented in the Experimental Procedures. Among hundreds of primary hits cross-screened for reversibility, dose-responsive potency, and selectivity with respect to inhibition of medium- and long-chain coenzyme-A-dependent acyl-CoA synthetase enzymes, a small molecule quinoxaline having an IC_{50} of $\sim 0.6 \mu\text{M}$ in biochemical assays, and $\sim 5 \mu\text{M}$ in its ability to inhibit cellular [^{14}C]acetate uptake into both lipids and histones (Figures 1E

(D) MEFs of the indicated genotypes were assayed for their ability to utilize [^{14}C]acetate for histone acetylation. Note the $^{-/-}$ MEFs show very little [^{14}C]acetate incorporation into histones (mean \pm SD, $n = 3$).

(E) ACSS2 is druggable. A high-throughput screen was conducted to identify small molecule inhibitors of the human ACSS2 enzyme (Experimental Procedures). The structure of one of the most potent and specific inhibitors, 1-(2,3-di(thiophen-2-yl)quinoxalin-6-yl)-3-(2-methoxyethyl)urea, is shown. This quinoxaline compound inhibited the ability of HepG2 cells to incorporate [^{14}C]acetate into lipids with $\text{IC}_{50} = 6.8 \mu\text{M}$ (mean \pm SEM, $n = 3$).

(F) The quinoxaline was also able to inhibit HepG2 utilization of [^{14}C]acetate for histone acetylation with $\text{IC}_{50} = 5.5 \mu\text{M}$ (mean \pm SEM, $n = 3$).

(G) Knockdown of ACSS2 in cancer cell lines significantly reduces acetate incorporation into lipids. [^{14}C]acetate incorporation into lipids was assayed in LL/2, PC3, U2OS, or Hep3B cancer cell lines harboring stable knockdown of ACSS2 or control (REN). Knockdown efficiency is shown in Figure S1. For each cell line, uptake amount was normalized against control (mean \pm SD, $n = 3$).

(H) Knockdown of ACSS2 in cancer cell lines significantly reduces acetate incorporation into histones. [^{14}C]acetate incorporation into histones was assayed in cancer cell lines harboring stable knockdown of ACSS2 or control (REN) (mean \pm SD, $n = 3$).

and 1F), emerged as one of the most favorable inhibitors of ACSS2. Moreover, this inhibitor did not reduce the residual acetate uptake observed in *ACSS2*^{-/-} MEFs (Figure S1). The fact that a selective chemical inhibitor of ACSS2 substantially inhibits acetate incorporation into both lipids and histones further confirms the importance of this enzyme for cellular uptake of acetate.

As a final test of the importance of ACSS2 for acetate uptake, small hairpin RNA (shRNA)-mediated attenuation of ACSS2 was tested on four different cancer cell lines, including LL/2, PC3, U2OS, and Hep3B (Figure S1). In all four cancer cell lines, attenuation of ACSS2 led to a significant impediment to [¹⁴C]acetate uptake into both lipids and histones (Figures 1G and 1H). The combination of shRNA-mediated mRNA knockdown approaches, studies of MEFs selectively mutated at the *ACSS2* locus, and the identification of a selective chemical inhibitor of ACSS2 provide strong evidence that this particular enzyme is primarily responsible for allowing mammalian cells to convert acetate into acetyl-CoA for subsequent metabolic utilization.

Reduced Tumor Formation in *ACSS2*-Deficient Mice

As a means of testing the contribution of ACSS2 to tumorigenesis, we introduced the *ACSS2*-null allele into a strain of genetically engineered mice that develop liver cancer due to expression of a liver-specific, doxycycline (dox)-regulated transgene encoding the SV40 early region (ApoE-rtTA_{M2}:TRE2-TAg; herein referred to as TAg). Notably, adult *ACSS2*-null mice display no overt phenotypic deficits and both sexes are fertile. Cohorts of male and female TAg mice with and without *ACSS2* were generated as described in Experimental Procedures and provided with drinking water supplemented with 10 µg/ml doxycycline for 42–45 days (Table S1). This regimen promotes reproducible and robust multifocal tumor development in a background of hepatic hyperplasia (Comerford et al., 2012). Post-sacrifice, livers were scored for tumor development using a nonlinear tumor-burden scale based on the number and size of visible tumors on the surface of the liver, percentage of liver/body weight, and the relative amount of tumor-free liver as described in detail (Table S2).

After 42–45 days of dox treatment, livers of *ACSS2*^{+/+}:TAg mice were covered with small- to medium-sized tumors reflecting multifocal tumor growth in response to sustained expression of the SV40 large T and small t (LT/st) oncoproteins. Twenty of 21 mice (95%) exhibited tumor burden scores of 8 or greater (Figure 2A). By contrast, only 48% of *ACSS2*^{-/-}:TAg mice exhibited tumor growth of an equivalent magnitude, with almost half (41%) receiving scores of 6 or lower. The statistical significance in the overall reduction in mean tumor burden score from 9.4 to 6.8 in mice with and without *ACSS2* achieved a calculated p value of 0.0002 (Figure 2A). Given that tumor burdens within the 9–10 range and 6–7 range reflected the presence of 100–200 tumors (too many to accurately count) or 20–50 tumors, respectively, these data indicate that *ACSS2* deficiency reduces the absolute number of liver tumors by at least 4-fold.

Having shown that loss of *ACSS2* correlates with reduced tumor burdens in the TAg liver cancer model, we next asked what impact of loss of *ACSS2* might have on the development of tumors driven by c-Myc overexpression and loss of *PTEN*,

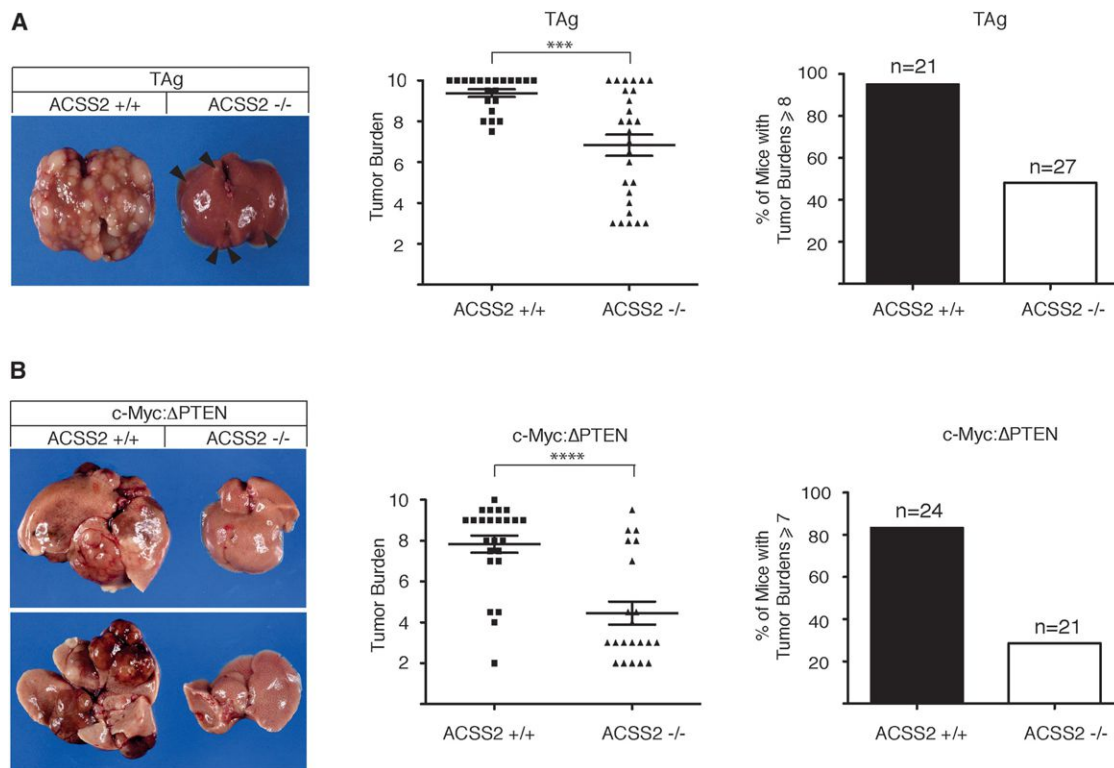
both of which are associated with human hepatocarcinogenesis (Kawate et al., 1999; Peng et al., 1993; Yao et al., 1999). In this genetically engineered mouse model, liver-specific expression of c-Myc (Sandgren et al., 1989), in conjunction with hepatic deletion of the *PTEN* tumor suppressor (*PTEN*^{lox/lox}) (Lesche et al., 2002), promotes the development of at least two large tumors (>10 mm) per liver in a background of hepatomegaly in 90% of mice at 6–7 months of age (S.A.C. and R.E.H., unpublished data). This second liver cancer model, herein referred to as c-Myc:Δ*PTEN*, constitutes a more stochastic and clinically relevant model of hepatic tumor development than the SV40 TAg model. Cohorts of male and female c-Myc:Δ*PTEN* mice with and without *ACSS2* were generated as described in Experimental Procedures and sacrificed at 6–7 months of age. Livers were scored for tumor development using a scale that differed from the one used to score the TAg tumors to accommodate the more stochastic pattern of tumor development in these mice (Table S2).

At 6–7 months of age, 20/24 (83%) of *ACSS2*^{+/+}:cMyc:Δ*PTEN* mice had livers with at least ten visible tumors (2–15 mm in size), resulting in assignment of tumor burden scores of 7 or greater (Figure 2B). This was in direct contrast to cohorts of age-matched *ACSS2*^{-/-}:c-Myc:Δ*PTEN* mice in which only 6/21 mice (29%) had tumor burdens of 7 or greater. Indeed, *ACSS2* deficiency not only reduced the mean tumor burden score from 7.8 to 4.7 (p < 0.0001), but inhibited tumor growth to such an extent that 71% of *ACSS2*-deficient mice had livers with fewer than ten small tumors, or no tumors at all (Figure 2B).

Immunohistochemical Studies of *ACSS2* Expression in Hepatocellular Tumors in Mice

Our results showing that the loss of *ACSS2* reduced liver tumor development in two different mouse models suggested that a subset of tumors might be dependent on *ACSS2* for growth and therefore express high levels of *ACSS2*. To determine if this was the case, we surveyed the distribution of *ACSS2* protein expression in normal mouse liver and in tumor-bearing livers from both cancer models by performing immunohistochemistry (IHC) with an *ACSS2*-specific antibody. In livers of WT mice, *ACSS2* was highly expressed in the cytoplasm and nucleus of hepatocytes in periportal zones 1 and 2 with virtually no expression evident in centrilobular zone 3 or in bile ducts (Figure 3A). Moreover, livers of *ACSS2*-null mice were devoid of any immunoreactivity indicating absolute specificity of the antibody for *ACSS2* and lack of any cross-reactivity with *ACSS1*, *ACSS3*, or other acyl-CoA synthetase enzymes (Figure 3A).

The expression of *ACSS2* in tumors from the TAg and c-Myc:Δ*PTEN* mice showed significant heterogeneity both between and within individual tumors (Figure 3A). Approximately 56% of TAg and 75% of c-Myc:Δ*PTEN* tumors contained *ACSS2*-positive cells (Table S3), whereas the remainder contained none. Given that *ACSS2*-positive tumors varied with respect to the proportion of *ACSS2*-expressing cells, a scoring method was devised to classify tumors as *ACSS2*^{HIGH}, *MED*, or *LOW* according to the percentage of *ACSS2*-positive cells within each tumor (see Experimental Procedures). Using this method, we determined that approximately half of the *ACSS2*-positive tumors in each tumor model (53% of TAg and 58% of c-Myc:Δ*PTEN*



tumors) were either ACSS2^{HIGH} or ACSS2^{MED}, whereas the remainder were ACSS2^{LOW} (Table S3). In several tumors, focal areas of prominent nuclear ACSS2 staining were seen, suggesting that enhanced translocation of ACSS2 from the cytoplasm to the nucleus might be associated with tumorigenesis, although this was not a universal finding. Importantly, evaluation of ACSS2 expression in c-Myc:ΔPTEN livers, where tumors were juxtaposed to areas of liver in which normal hepatic zonation had been maintained, showed that the level of ACSS2 expressed in ACSS2^{HIGH} tumors was comparable to that found in normal periportal hepatocytes, suggesting that ACSS2 expression was maintained, rather than induced, in a subset of tumors, and was lost in others. As expected, tumor-bearing livers from ACSS2-deficient TAg or c-Myc:ΔPTEN mice were devoid of any immunoreactivity.

Having shown that the loss of ACSS2 reduced tumor incidence in both liver cancer models, we next determined if ACSS2 deficiency altered the tumor spectrum. Comparative

histological analysis of hematoxylin-and-eosin-stained sections of tumor-bearing livers of ACSS2^{+/+} and ACSS2^{-/-} TAg and c-Myc:ΔPTEN mice showed that ACSS2 deficiency increased the proportion of well-differentiated HCCs at the expense of moderately and poorly differentiated HCCs and other tumor types in both models (Figure S2), suggesting that ACSS2 drives the emergence of more aggressive hepatic tumors in the mouse.

Immunohistochemical Studies of ACSS2 Expression in Human Tumors

Clinical imaging studies have observed avid [¹¹C]acetate uptake in a wide spectrum of human cancers, including prostate, liver, lung, and brain tumors (Ho et al., 2003; Nomori et al., 2008; Oyama et al., 2002; Tsuchida et al., 2008). It has been proposed that acetate uptake in these tumors might reflect expression of the acetyl-CoA synthetase enzymes, including ACSS2 (Yoshii et al., 2009b; Yun et al., 2009). In normal adult mice, ACSS2 mRNA is most abundantly expressed in liver

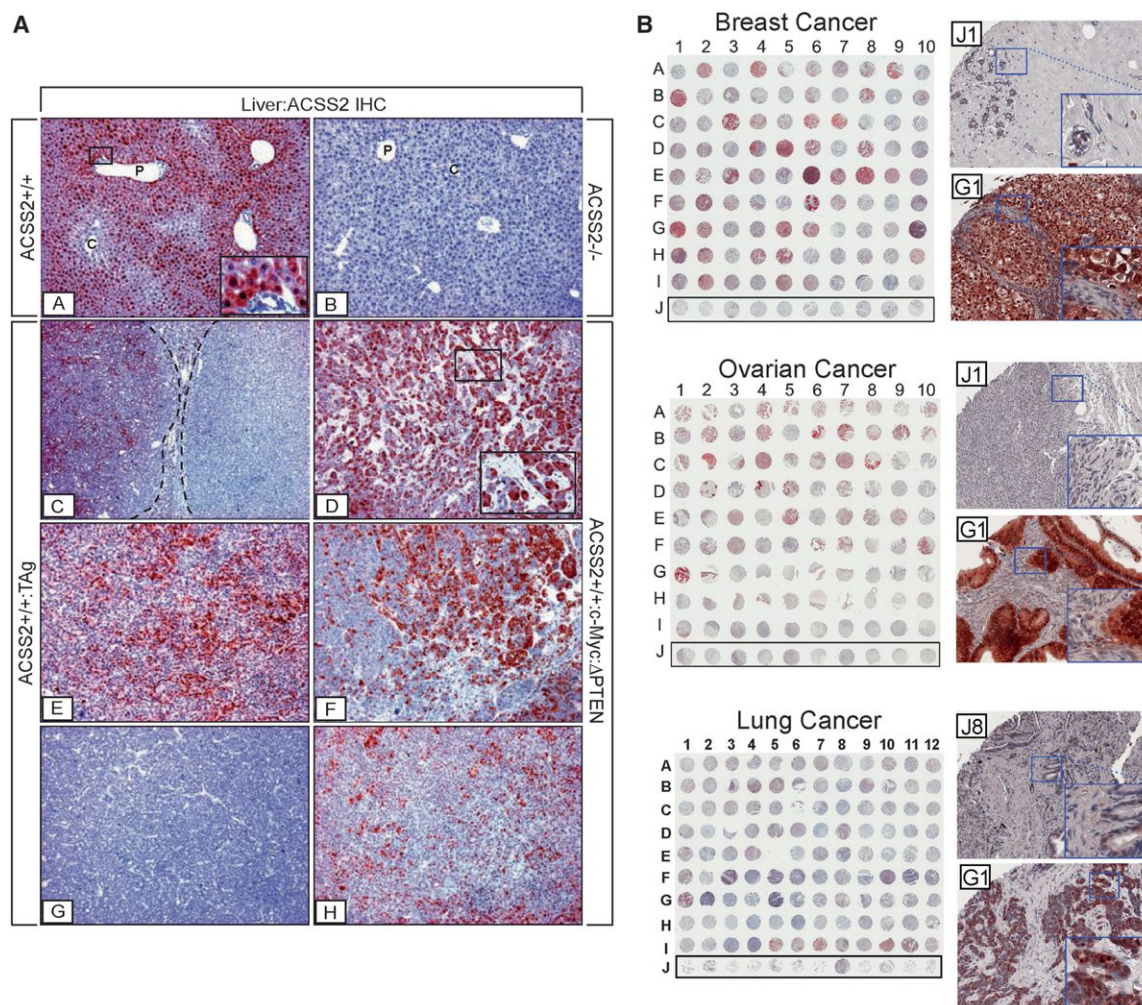


Figure 3. Immunohistochemical Analysis of ACSS2 Expression in Tumors

(A) Expression patterns of ACSS2 protein in normal mouse liver and liver tumors. Insets: [A] ACSS2 IHC on WT liver showing regional expression of ACSS2 across the lobule. ACSS2 is localized to the nucleus/cytoplasm in zone 1 and 2 hepatocytes but is largely absent from zone 3 hepatocytes and biliary cells. P, portal vein; C, central vein. [B] Complete absence of ACSS2 expression in liver of an ACSS2^{-/-} mouse. [C, E, and G] Variability of ACSS2 expression in tumor-laden livers of ApoE-rTA:TRE2-TAg mice provided with 10 μ g/ml doxycycline for 42–45 days. [D, F, and H] ACSS2 expression in c-Myc: Δ PTEN tumors. AEC chromagen (red); hematoxylin counterstain (blue). [A] $\times 60$ mag, higher mag inset $\times 366$; [B] $\times 60$ mag; [C] $\times 36$ mag; [D] $\times 60$ mag, higher mag inset $\times 244$; [E]–[H] $\times 60$ mag.

(B) Survey of ACSS2 expression in human breast, ovarian, and lung tumors. IHC was carried out to assess ACSS2 protein expression in a panel of human breast, ovarian, and lung tumor sections in tissue microarray format. Row J denotes samples of normal, noncancerous tissue. For each category of tumor, a higher magnification image of one representative example of staining in normal tissue, and one representative example of staining in a high ACSS2-expressing tumor, are shown to the right. Scoring of ACSS2 expression in these TMA samples is indicated in Table S5.

and kidney, with lower levels present in the heart, brain, and testis (Luong et al., 2000). We next surveyed the extent of ACSS2 protein expression in a variety of human tumor samples by immunohistochemical staining of commercially available tumor tissue microarrays. Significant ACSS2 expression was observed in a substantial number of human breast, ovarian, and lung tumor samples (Figure 3B). Normal, noncancerous samples of the tissues of origin exhibited little to no ACSS2 protein at all.

We then surveyed ACSS2 expression in a well-annotated set of human triple-negative breast cancers (Figures 4A and S3).

Immunohistochemical analysis was available for 154 cases after discounting loss of cores due to sectioning and IHC processing. Expression was stratified into three categories for association with overall survival. High expression of ACSS2 (H-score 120–300) was associated with shorter overall survival compared to ACSS2-negative cases ($p = 0.03$) (Figure 4B). However, the observed correlation between ACSS2 expression and decreased survival does not necessarily imply dependency. Expression of ACSS2 was observed predominantly in the cytoplasm with some cases exhibiting significant nuclear staining. Taken together, these observations strongly suggest that

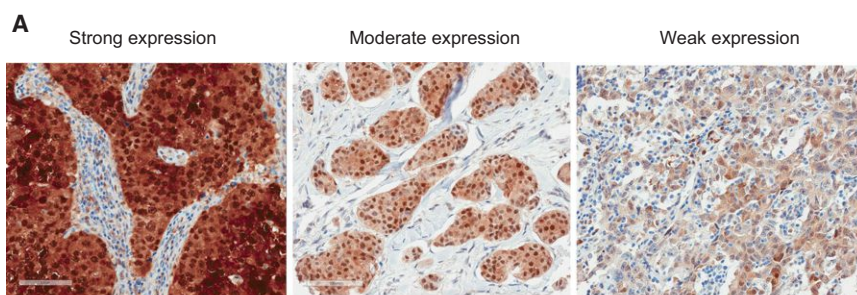
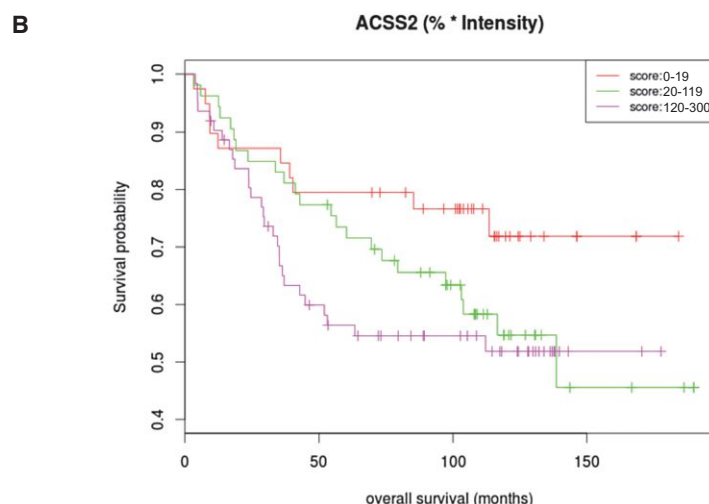


Figure 4. ACSS2 Expression in Triple-Negative Breast Cancers

(A) Representative examples of high (H-score 120–300), moderate (H-score 20–119), and low (H-score 0–19) ACSS2 expression in triple-negative breast cancers. The complete TMA staining is shown in Figure S3.

(B) High expression of ACSS2 was associated with poorer overall survival ($p = 0.03$).



ACSS2 is expressed to a significant extent in particular tumor types, including triple-negative breast cancers.

[¹¹C]Acetate PET Imaging of Hepatocellular Tumors in Mice

To directly test whether [¹¹C]acetate uptake correlated with ACSS2 expression in the mouse models of liver cancer, we performed [¹¹C]acetate PET imaging of tumors developing in the c-Myc:ΔPTEN and TAG mice. We observed the expected biodistribution for [¹¹C]acetate in WT mice, with modest uptake in normal liver. In contrast, mice bearing liver tumors exhibited greater uptake in the liver, with focal regions of significantly greater [¹¹C]acetate uptake (Figures 5A–5D). Tracer pharmacokinetics appeared to reach a state of equilibrium between 25 and 40 min after radiotracer administration, as depicted in the time-activity curves (TAC) (Figure 5E). During this time interval, radiotracer concentration within the tumor region-of-interest (ROI) was nearly three times greater in tumors arising in the c-Myc:ΔPTEN mice compared to uptake in normal liver. Presence of tumor was confirmed histologically and exhibited significant staining of ACSS2 (Figure 5F). The imaged tumor in the TAG mouse exhibited only modestly higher [¹¹C]acetate uptake, which correlated with lower expression of ACSS2. Moreover, several [¹¹C]acetate-negative tumors arising in the TAG mice showed virtually no expression of ACSS2 protein at all, and [¹¹C]acetate uptake was lower in liver from an ACSS2^{−/−} mouse

compared to WT (Figure S4). Collectively, these observations demonstrate a strong correlation between [¹¹C]acetate uptake and ACSS2 expression.

DISCUSSION

Here, we describe experiments indicating that certain tumors may have acquired a dependency on acetate as a carbon source for the production of acetyl-CoA. This conclusion is, for many reasons, perplexing. Unlike the low millimolar levels of blood glucose, which is a considerably more energetic carbon source than acetate, blood level measurements for acetate are estimated at ~20–50 μM (Psychogios et al., 2011; Tollinger et al., 1979). How could an energy-diminished metabolite, relative to glucose, feed tumors when its abundance in blood

is two orders of magnitude lower than glucose? Before offering possible explanations to this conundrum, it may be of value to compare the evolution of human tumors to the work of Richard Lenski and colleagues on bacterial evolution (Blount et al., 2012). By continually growing cultures of *E. coli* for upward of 25 years, Lenski and colleagues have evolved derivatives that outcompete the starting bacterial cells via enhanced cell growth rates in the range of 70%. Some of Lenski's cultures were found to have achieved enhanced growth via the induction of pathways facilitating the catabolism of citrate. We offer the hypothesis that during the evolution of certain tumors, a Lenski-like adaptation may have taken place—not to citrate dependence, but to acetate dependence.

In support of this idea, we offer the following considerations. [¹¹C]acetate PET imaging studies have shown that many human cancers avidly take up acetate, and confirmatory [¹³C]acetate NMR tracer studies also reveal selective acetate uptake in human tumors (Mashimo et al., 2014). Moreover, substantive ACSS2 immunohistochemical staining of both hepatic tumors in mice and numerous tumors from humans reveals strong expression of the enzyme (Figures 3 and 4). Indeed, we observe a strong correlation between ACSS2 expression and uptake of [¹¹C]acetate into liver tumors in mice (Figure 5). As shown here, for mammalian cells ACSS2 is the most important member of this enzyme family for uptake of acetate and its conversion into lipids and acetylated histones (Figure 1). Last, animals bearing

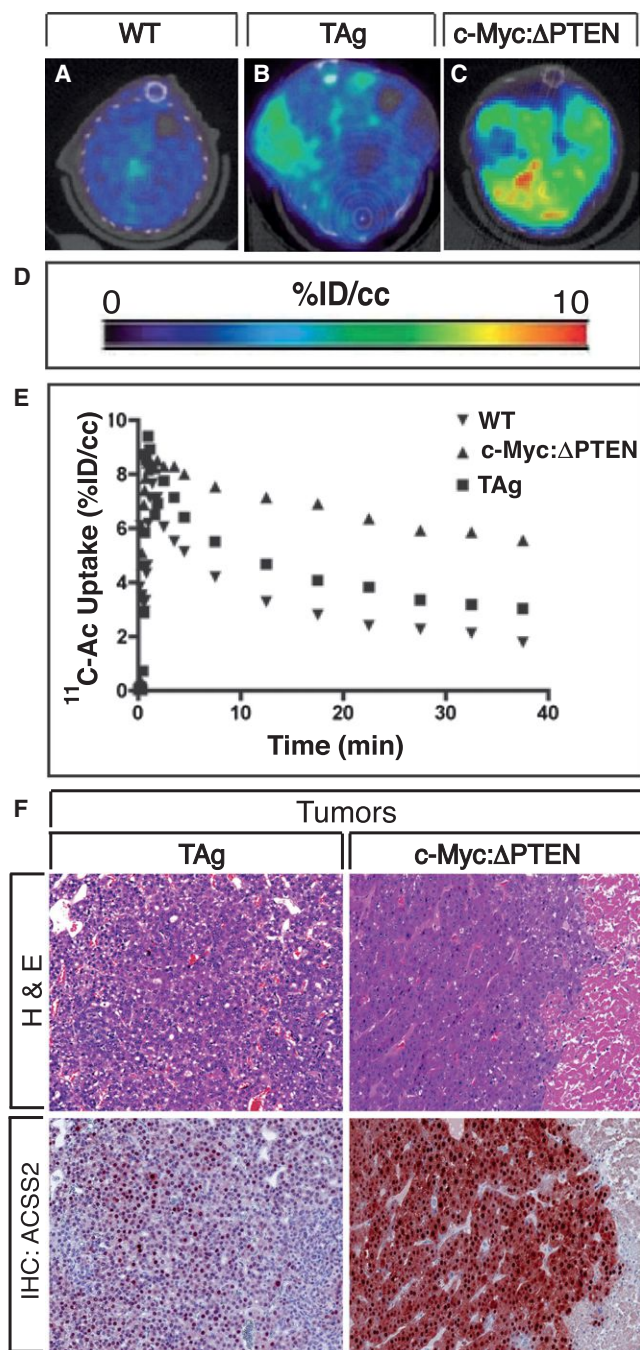


Figure 5. ^{11}C Acetate Uptake Correlates with ACSS2 Expression in HCC Driven by TAg or Combined Overexpression of c-Myc and Loss of PTEN

(A–C) Combined ^{11}C acetate PET/CT of WT (A), TAg (B), and c-Myc:ΔPTEN (C) mice (transverse views).

(D) Scale (%ID/cc) of avidity of ^{11}C acetate uptake.

(E) Time activity curve (TAC) of 40 min dynamic PET scan of ^{11}C acetate uptake in WT, TAg, and c-Myc:ΔPTEN mice shown in (A)–(C).

(F) Photomicrographs ($\times 125$ magnification) of H&E (top) and corresponding ACSS2 IHC (bottom) of tumors from TAg (left hand panels) and c-Myc:ΔPTEN (right hand panels) mice shown in (B) and (C). ACSS2 staining (red/brown color) is predominantly nuclear in the poorly differentiated tumor

targeted mutations in both alleles of the ACSS2 gene show a substantially diminished tumor burden in two genetic models of cancer (Figure 2). We simplistically conclude that certain tumors are able to avidly capture acetate to help fuel their growth and survival (Figure 6). If normal cells and tissues are far less competent in acquiring and retaining acetate, then tumors may be able to survive and grow by avid consumption of this suboptimal hydrocarbon fuel.

Studies of prototrophic yeast grown continuously in a chemostat offer a framework for thinking about how mammalian tissues might share acetate. When grown at high density under continuous, glucose-limitation conditions, yeast cells enter into a robust metabolic cycle in which a population of yeast cells synchronously oscillates (Tu et al., 2005). During glycolytic growth, cells consume glucose and ferment the carbon source all the way to ethanol. By contrast, during respiratory growth, ethanol is retrieved as a carbon source via its sequential anabolic conversion into acetaldehyde, acetate, and acetyl-CoA. Indeed, the periodic secretion and uptake of ethanol and acetate can be observed as a function of the yeast metabolic cycle (Tu et al., 2005), revealing that these simple products of glycolysis can be secreted and then shared among subpopulations of the community.

We offer the speculative idea that acetate is liberally moved in and out of mammalian cells and tissues in a dynamic fashion. Because acetate is not typically thought to be a physiologically significant carbon source in mammals, what are the possible sources of acetate? The half-life of acetyl modifications on histones has been measured to be on the order of only ~ 2 – 3 min as a result of the opposing actions of histone acetyltransferase and deacetylase enzymes (Jackson et al., 1975; Waterborg, 2002). Histone proteins are among the most abundant in cells, and several dozen sites within each histone octamer are subject to modification by acetylation (Shahbazian and Grunstein, 2007). Considering the extremely short half-life of acetylation modifications, substantial quantities of acetate could be continuously released from histone tails simply as a result of deacetylation. Moreover, if certain tumor cells liberally secrete lactate via the truncation in glycolysis discovered by Otto Warburg, neighboring normal cells could convert lactate back into pyruvate, thereby yielding a possible source of acetyl-CoA. Acetyl-CoA hydrolase or thioesterase enzymes could rebalance metabolism such that acetate could be liberated from normal cells as a hydrocarbon source for rogue consumption by tumor cells. In other words, if a tumor cell were hard-wired to exist in an acetate-capturing or fixing state, it might significantly outcompete the rest of the human body in its ability to feed upon a suboptimal carbon source present at relatively low concentrations.

At the cost of a single ATP, one acetate molecule can be retrieved to produce acetyl-CoA for use in the synthesis of fatty acids or sterols, for the acetylation of histones, or for further oxidation via the TCA cycle to generate an additional ~ 12 units of ATP (Figure 6). Acetate can also be used for the synthesis of the amino acid glutamate (Mashimo et al., 2014). Unlike yeast,

in the TAg mouse (left) but is both nuclear and cytoplasmic in the moderately differentiated trabecular HCC in the c-Myc:ΔPTEN mouse (right). See also Figure S4.

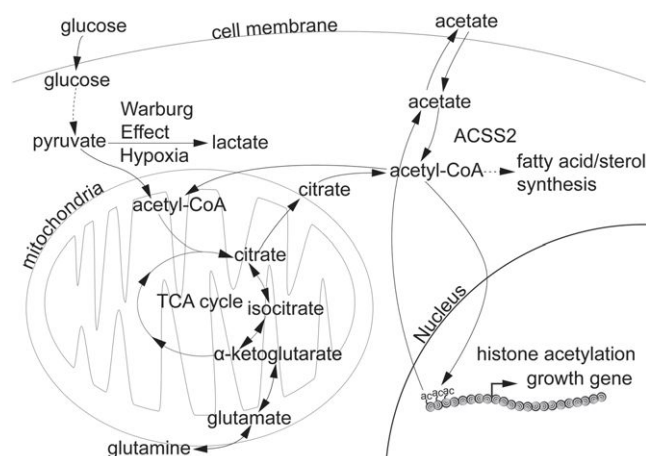


Figure 6. Acetate Is a Critical Source of Acetyl-CoA for Certain Tumors

Schematic diagram depicting pathways that synthesize and consume acetyl-CoA in mammalian cells. Hypoxic or highly glycolytic cancer cells preferentially shunt pyruvate to lactate, instead of to acetyl-CoA via the pyruvate dehydrogenase complex, raising the question of how such cells obtain sufficient quantities of acetyl-CoA. Among numerous metabolic functions, acetyl-CoA is used for fatty acid/sterol synthesis, histone acetylation, the synthesis of glutamate (Mashimo et al., 2014), or further oxidation via the TCA cycle for ATP synthesis. Glutamine can reportedly serve as a source of acetyl-CoA in cell culture studies. Acetate is an overlooked source of acetyl-CoA and can be produced as a result of histone or protein deacetylation, or from the action of acetyl-CoA thioesterase/hydrolase enzymes. The nucleocytosolic acetyl-CoA synthetase enzyme ACSS2 enables the recapture of acetate to acetyl-CoA that can subsequently be utilized for the indicated metabolic processes, all of which are expected to support tumor growth or survival.

mammals lack a glyoxylate cycle and therefore cannot utilize acetate for the synthesis of glucose. Nonetheless, the ability to recoup acetate for any of these other purposes stands to promote tumor cell growth or survivability in the face of nutritionally challenging or hypoxic microenvironments. In contrast, acetate might be irrelevant for nutritionally replete cells or tissues. Given ample supplies of glucose, coupled with the ability to convert glucose into acetyl-CoA via citrate spilling out of mitochondria, normal cells or tissues might actually represent net producers of acetate, rather than net consumers.

Were it indeed the case that certain tumors acquire, via expression of ACSS2, a strict dependency on acetate for their growth or survival, then selective inhibitors of this nonessential enzyme might represent an unusually ripe opportunity for the development of new anticancer therapeutics. By screening a library of 200,000 drug-like chemicals, we have discovered compounds that are capable of selectively inhibiting ACSS2 relative to other acyl-CoA synthetase enzymes that ligate coenzyme A onto medium or long chain fatty acids. It is hoped that these chemical inhibitors of ACSS2 can be optimized for potency, selectivity, and pharmacological properties. If the normal human cells and tissues are not heavily reliant on the activity of the ACSS2 enzyme, it is possible that such agents might inhibit the growth of ACSS2-expressing tumors with a favorable therapeutic window.

EXPERIMENTAL PROCEDURES

Construction of Stable Knockdown Cell Lines

Cells were seeded in 60 mm dishes to 30% confluence the day before infection. Fresh medium (1 ml) and retrovirus (1 ml; treated with 8 μ g/ml polybrene) carrying shRNA against ACSS1, ACSS2, ACSS3, or REN (Renilla luciferase) as control were added during virus infection. Puromycin was added at 2 μ g/ml to the medium for selection at 48–60 hr after infection. Medium was replaced with fresh medium supplemented with puromycin every 1 or 2 days. Cells were split as they became confluent. Puromycin was maintained in the medium for selection up to 10 days. The following are the sequences of shRNA used.

shREN: AGGAATTATAATGCTTATCTA
 shACSS1: CCAGTTAAATGTCTGTCAA
 shACSS2: CAGGATTGATGACATGCTCAA
 shACSS3: GCCGTTGATCGTCATATTGAA

[14 C]Acetate Incorporation into Lipid Fractions

Cells were grown in 12-well plates to 70%–80% confluence. Cells were treated with 1 μ Ci/ml sodium [14 C]acetate (PerkinElmer) for 6 hr. After two washes with ice-cold PBS, cells were lysed with 0.6 ml MeOH solution (MeOH:H₂O = 2.5:1). CHCl₃ (0.4 ml) was added to lysate and mixed by vortexing for 30 s. Lysates were then centrifuged for 5 min at 1,000 rpm for phase separation. The lipid soluble fraction was collected as the lower layer. Fractions were counted for radioactivity.

[14 C]Acetate Incorporation into Histones

Cells were grown in 6-well plates to 70%–80% confluence. Cells were treated with 1 μ Ci/ml [14 C]acetate for 6 hr. After two washes with ice-cold PBS, cells were lysed in hypotonic buffer (10 mM Tris-Cl [pH 8.0], 1 mM KCl, 1.5 mM MgCl₂, 1 mM DTT) with protease inhibitors and then subjected to one freeze/thaw cycle at -20° C. Lysates were centrifuged at 4° C 10,000 \times g for 10 min. Supernatants were discarded, and the pellets were resuspended in 400 μ l 0.4 N H₂SO₄ and vortexed until pellets were dissolved. The lysates were rotated at 4° C overnight and then centrifuged at 4° C 16,000 \times g for 10 min. Supernatants were taken and counted for radioactivity.

Mouse Tumorigenesis Experiments and Scoring of Tumor Burdens

Generation of mice and liver tumorigenesis experiments are described in detail in the Supplemental Information. All experiments were approved by the University of Texas Southwestern Medical Center Institutional Animal Care and Use Committee. Tumor burden scores were assigned on the basis of the number and size of tumors, the magnitude of hepatic hyperplasia, and the amount of tumor-free liver remaining after 42–45 days of dox treatment (TAg model) or at 6–7 months of age (c-Myc: Δ PTEN model). A detailed description of criteria used to assign tumor burden scores in each model is provided in Table S2. Scores were assigned by S.A.C. or R.E.H.

Scoring of ACSS2 Expression in Tumors

Tumor scoring for ACSS2 expression was conducted by performing IHC with the ACSS2-specific antibody on formalin fixed paraffin-embedded sections of liver from TAg mice provided with dox for 42–45 days and from 6- to 7-month-old c-Myc: Δ PTEN mice. Each slide contained sections cut from a block containing a minimum of six randomly selected pieces of liver collected at the time of sacrifice. A total of 80 tumors from the TAg mice and 32 from the c-Myc: Δ PTEN mice were surveyed. The average number of TAg-induced tumors surveyed/slide was 20 (range; 15–25), whereas the average number of c-Myc: Δ PTEN-driven tumors/slide was eight (range; seven to ten). Tumors were scored on the basis of the number of ACSS2-positive cells within each tumor. Tumors were designated ACSS2^{HIGH}, ACSS2^{MED}, or ACSS2^{LOW} if they contained >60%, 30%–60%, or <30% ACSS2-positive cells, respectively. Tumors that did not contain any visible ACSS2-positive cells were designated ACSS2^{NEG}. Tumor-specific expression of ACSS2 was independently scored by S.A.C. and Z.H.

Human Triple-Negative Breast Cancer Tissue Microarray

Tissue microarray with 168 triple-negative breast carcinoma (TNBC) cases was employed to evaluate ACSS2 protein expression by immunohistochemistry. Tissues present on the array were obtained with institutional review board approval. TNBC clinical designation was defined as the absence of staining for estrogen receptor, progesterone receptor, and *HER2/neu*. Clinical and pathologic variables were determined following well-established criteria. All invasive carcinomas were graded according to the method described by Elston and Ellis (1991). The tissue microarrays (TMAs) were constructed as previously described using a tissue arrayer (Beecher Instruments). Two tissue cores (0.6 mm diameter) were sampled from each block and transferred to the recipient TMA block.

All ACSS2-stained slides were scanned using the ScanScope System (Aperio Technologies) and viewed using ImageScope software (Aperio). Pathologist with subspecialty training in breast pathology (A.K.W.) assured that areas selected for automated image analysis represented tumor. Staining intensity was graded with: none as 0, weak as 1+, moderate as 2+, and marked as 3+. The percentage of stained tumor cells was recorded. H-score was calculated as previously described using the following formula ($1 \times \% \text{ of weak staining} + 2 \times \% \text{ moderate staining} + 3 \times \% \text{ marked staining}$). Duplicates of cores were scored separately, and the highest-yielding H-score among the three constructs was selected.

Data and associations were assessed using the R-statistical package. Kaplan-Meier survival curves were used to estimate overall survival, which was defined as the time from date of surgery to date of death from TNBC or to the last follow-up date for censored cases. The log-rank test was performed to compare survival between groups. A p value < 0.05 was considered significant.

High-Throughput Screen to Identify Inhibitors of Human ACSS2

Human ACSS2 enzyme was expressed in insect cells using the Invitrogen Bac to Bac expression system. Acetyl-CoA (C4780) and pyrophosphatase (I1643) were purchased from Sigma. CellTiter-Glo was purchased from Promega. EnzChek Pyrophosphate Assay Kit (E6645), which includes purine nucleotide phosphorylase (PNP) and MESG, was purchased from Life Technologies.

ACSS2 converts ATP, acetate, CoA into AMP, pyrophosphate, and acetyl-CoA. In our assay, the reaction was monitored by the disappearance of ATP using a luciferase-based CellTiter-Glo reagent. The screen was carried out using 384-well plates. A typical reaction mixture contained 0.1 mM CoA, 0.02 mM ATP, 10 mM DTT, 0.1 unit/ml pyrophosphatase, 1 μ g/ml ACSS2, 0.1 mg/ml BSA, 150 mM NaCl, 1 mM MgCl_2 , and 50 mM HEPES (pH 7.5). The reaction was initiated by the addition of 5 μ l of 25 mM sodium acetate to each well containing 20 μ l reaction mixture. The plates were then spun for 1 min at 1,500 $\times g$ and incubated at 37°C. When the reaction reached 70% completion, which typically occurred after 2.5–5 hr, 25 μ l CellTiter-Glo reagent was added to each well. The plates were shaken for 2 min, and luminescence was read on a PerkinElmer Envision plate reader. The presence of an inhibitor slowed down the consumption of ATP and was revealed by high luminescence reading on the plate reader. We identified 1,152 hits in the primary screen of 220K compounds in our chemical library at UTSW. By removing redundant hits in the large structure groups, we cherry-picked 926 hits for further confirmation using the same assay but carried out in triplicate. We were able to confirm 118 of the 926 hits. To test the specificity of the inhibitors, we carried out a counterscreen against ACSF2 (medium chain fatty acid-CoA synthetase) and ACSL5 (long chain fatty acid-CoA synthetase). We were able to identify 62 compounds as specific inhibitors against ACSS2 based on their nonactivity against ACSF2 and ACSL5. One of these most potent and specific inhibitors N-(2,3-di-2-thienyl-6-quinoxaliny)-N'-(2-methoxyethyl)urea was purchased from ChemBridge and re-tested (Figures 1E and 1F).

Biochemical Measurement of IC_{50} of Inhibitors

The measurement of initial rate is ideally carried out when the reaction extent is less than 15%. The variation in the luciferase assay used in the HTS screen is typically around 10%. For this reason, we adopted a different assay for the measurement of reaction rate in the determination of IC_{50} of ACSS2 inhibitors.

Instead of monitoring the consumption of ATP, we measured generation of pyrophosphate in a coupled enzymatic assay. The reaction mixture contained pyrophosphatase, which converts pyrophosphate to phosphate, and purine nucleotide phosphorylase (PNP), which couples phosphate with a nucleotide analog (MESG), resulting in a compound that absorbs light at 360 nm. A typical reaction mixture includes 50 μ M CoA, 100 μ M ATP, 5 mM sodium acetate, 0.2 mM MESG, 0.2% BSA, 20 nM ACSS2, 0.4 μ M PNP, 1 unit/ml of yeast pyrophosphatase, 1 mM MgCl_2 , 150 mM NaCl, and 50 mM HEPES (pH 7.5). The reaction rates were computed by the increase of absorption at 360 nm for a range of inhibitor concentrations. Curve fitting (rate versus inhibitor concentration) in PRISM yielded IC_{50} .

Construction of ACSS2 KO mice, mouse breeding and genotyping procedures, isolation of MEFs, and [^{14}C]acetate PET methods are described in Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.020>.

AUTHOR CONTRIBUTIONS

This study was conceived by B.P.T. and S.L.M. S.A.C. and R.E.H. conducted all of the in vivo liver tumorigenesis experiments and isolated the MEFs; Z.H., Y.W., and L.C. performed the cell line experiments; X.D. conducted the high-throughput screen; Z.H. and S.A.C. performed the IHC experiments; A.K.W. provided the collection of human triple-negative breast cancers; M.N.T., A.F., and H.C.M. performed the [^{14}C]acetate PET imaging experiments; R.E.H., H.W., and J.D.H. constructed the ACSS2 KO mice; S.A.C., R.E.H., Z.H., L.C., B.P.T., and S.L.M. designed the experiments; and S.A.C., Z.H., B.P.T., and S.L.M. wrote the manuscript with the assistance and approval of all authors.

ACKNOWLEDGMENTS

We thank B. Posner and the staff of the UTSW High-Throughput Screening Core Facility for assistance with the ACSS2 inhibitor screen; V. Vemireddy, E. Maher, R. Bachoo, S. Altschuler, and L. Wu for use of slide scanners; and E. Maher, R. Bachoo, and D. Nijhawan for helpful discussions. This work was supported by a Damon Runyon-Rachleff Innovation Award, Packard Fellowship, and a CPRIT High-Risk/High-Impact grant to B.P.T., R01CA185169 to B.P.T. and S.L.M., unrestricted funds provided to S.L.M. by an anonymous donor, and Excellence in Education funds (EEF) from UTSW to R.E.H. S.L.M. is the founder and Chairman of the Scientific Advisory Board, X.D. is an employee, B.P.T., S.A.C., and R.E.H. are consultants, and S.L.M., B.P.T., S.A.C., X.D., R.E.H., and L.C. are shareholders of Peloton Therapeutics, Inc.

Received: March 18, 2014

Revised: May 30, 2014

Accepted: November 11, 2014

Published: December 18, 2014

REFERENCES

- Ariannur, P.S., Moffett, J.R., Madhavarao, C.N., Arun, P., Vishnu, N., Jacobowitz, D.M., Hallows, W.C., Denu, J.M., and Nambodiri, A.M. (2010). Nuclear-cytoplasmic localization of acetyl coenzyme a synthetase-1 in the rat brain. *J. Comp. Neurol.* 518, 2952–2977.
- Berg, P. (1956). Acyl adenylates; an enzymatic mechanism of acetate activation. *J. Biol. Chem.* 222, 991–1013.
- Blount, Z.D., Barrick, J.E., Davidson, C.J., and Lenski, R.E. (2012). Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489, 513–518.

- Cai, L., and Tu, B.P. (2011). On acetyl-CoA as a gauge of cellular metabolic state. *Cold Spring Harb. Symp. Quant. Biol.* 76, 195–202.
- Cai, L., Sutter, B.M., Li, B., and Tu, B.P. (2011). Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. *Mol. Cell* 42, 426–437.
- Comerford, S.A., Schultz, N., Hinnant, E.A., Klapproth, S., and Hammer, R.E. (2012). Comparative analysis of SV40 17kT and LT function in vivo demonstrates that LT's C-terminus re-programs hepatic gene expression and is necessary for tumorigenesis in the liver. *Oncogenesis* 1, e28.
- De Virgilio, C., Bürckert, N., Barth, G., Neuhaus, J.M., Boller, T., and Wiemken, A. (1992). Cloning and disruption of a gene required for growth on acetate but not on ethanol: the acetyl-coenzyme A synthetase gene of *Saccharomyces cerevisiae*. *Yeast* 8, 1043–1051.
- Elston, C.W., and Ellis, I.O. (1991). Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 19, 403–410.
- Fujino, T., Kondo, J., Ishikawa, M., Morikawa, K., and Yamamoto, T.T. (2001). Acetyl-CoA synthetase 2, a mitochondrial matrix enzyme involved in the oxidation of acetate. *J. Biol. Chem.* 276, 11420–11426.
- Ho, C.L., Yu, S.C., and Yeung, D.W. (2003). 11C-acetate PET imaging in hepatocellular carcinoma and other liver masses. *J. Nucl. Med.* 44, 213–221.
- Jackson, V., Shires, A., Chalkley, R., and Granner, D.K. (1975). Studies on highly metabolically active acetylation and phosphorylation of histones. *J. Biol. Chem.* 250, 4856–4863.
- Jones, M.E., Lipmann, F., Hilz, H., and Lynen, F. (1953). On the enzymatic mechanism of Coenzyme A acetylation with adenosine triphosphate and acetate. *J. Am. Chem. Soc.* 75, 3285–3286.
- Kaelin, W.G., Jr., and McKnight, S.L. (2013). Influence of metabolism on epigenetics and disease. *Cell* 153, 56–69.
- Kawate, S., Fukusato, T., Ohwada, S., Watanuki, A., and Morishita, Y. (1999). Amplification of c-myc in hepatocellular carcinoma: correlation with clinicopathologic features, proliferative activity and p53 overexpression. *Oncology* 57, 157–163.
- Le, A., Lane, A.N., Hamaker, M., Bose, S., Gouw, A., Barbi, J., Tsukamoto, T., Rojas, C.J., Slusher, B.S., Zhang, H., et al. (2012). Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in B cells. *Cell Metab.* 15, 110–121.
- Leonardi, R., Subramanian, C., Jackowski, S., and Rock, C.O. (2012). Cancer-associated isocitrate dehydrogenase mutations inactivate NADPH-dependent reductive carboxylation. *J. Biol. Chem.* 287, 14615–14620.
- Lesche, R., Groszer, M., Gao, J., Wang, Y., Messing, A., Sun, H., Liu, X., and Wu, H. (2002). Cre/loxP-mediated inactivation of the murine Pten tumor suppressor gene. *Genesis* 32, 148–149.
- Lipmann, F., and Tuttle, L.C. (1945). The detection of activated carboxyl groups with hydroxylamine as interceptor. *J. Biol. Chem.* 161, 415.
- Luong, A., Hannah, V.C., Brown, M.S., and Goldstein, J.L. (2000). Molecular characterization of human acetyl-CoA synthetase, an enzyme regulated by sterol regulatory element-binding proteins. *J. Biol. Chem.* 275, 26458–26466.
- Marin-Valencia, I., Yang, C., Mashimo, T., Cho, S., Baek, H., Yang, X.L., Rajagopalan, K.N., Maddie, M., Vemireddy, V., Zhao, Z., et al. (2012). Analysis of tumor metabolism reveals mitochondrial glucose oxidation in genetically diverse human glioblastomas in the mouse brain in vivo. *Cell Metab.* 15, 827–837.
- Mashimo, T., Pichumani, K., Vemireddy, V., Hatanpaa, K.J., Singh, D.K., Sirasanagandla, S., Nannepaga, S., Piccirillo, S.G., Kovacs, Z., Fong, C., et al. (2014). Acetate is a bioenergetics substrate for human glioblastoma and brain metastases. *Cell* 159, this issue, 1603–1614.
- Metallo, C.M., Gameiro, P.A., Bell, E.L., Mattaini, K.R., Yang, J., Hiller, K., Jewell, C.M., Johnson, Z.R., Irvine, D.J., Guarente, L., et al. (2012). Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. *Nature* 481, 380–384.
- Mullen, A.R., Wheaton, W.W., Jin, E.S., Chen, P.H., Sullivan, L.B., Cheng, T., Yang, Y., Linehan, W.M., Chandel, N.S., and DeBerardinis, R.J. (2012). Reductive carboxylation supports growth in tumour cells with defective mitochondria. *Nature* 481, 385–388.
- Nomori, H., Shibata, H., Uno, K., Iyama, K., Honda, Y., Nakashima, R., Sakaguchi, K., Goya, T., Takanami, I., Koizumi, K., et al. (2008). 11C-Acetate can be used in place of 18F-fluorodeoxyglucose for positron emission tomography imaging of non-small cell lung cancer with higher sensitivity for well-differentiated adenocarcinoma. *Journal of thoracic oncology* 3, 1427–1432.
- Oyama, N., Akino, H., Kanamaru, H., Suzuki, Y., Muramoto, S., Yonekura, Y., Sadato, N., Yamamoto, K., and Okada, K. (2002). 11C-acetate PET imaging of prostate cancer. *J. Nucl. Med.* 43, 181–186.
- Peng, S.Y., Lai, P.L., and Hsu, H.C. (1993). Amplification of the c-myc gene in human hepatocellular carcinoma: biologic significance. *J. Formos. Med. Assoc.* 92, 866–870.
- Pérez-Chacón, G., Astudillo, A.M., Balgoma, D., Balboa, M.A., and Balsinde, J. (2009). Control of free arachidonic acid levels by phospholipases A2 and lysophospholipid acyltransferases. *Biochim. Biophys. Acta* 1791, 1103–1113.
- Psychogios, N., Hau, D.D., Peng, J., Guo, A.C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., et al. (2011). The human serum metabolome. *PLoS ONE* 6, e16957.
- Sandgren, E.P., Quaife, C.J., Pinkert, C.A., Palmiter, R.D., and Brinster, R.L. (1989). Oncogene-induced liver neoplasia in transgenic mice. *Oncogene* 4, 715–724.
- Shahbazian, M.D., and Grunstein, M. (2007). Functions of site-specific histone acetylation and deacetylation. *Annu. Rev. Biochem.* 76, 75–100.
- Shi, L., and Tu, B.P. (2013). Acetyl-CoA induces transcription of the key G1 cyclin CLN3 to promote entry into the cell division cycle in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 110, 7318–7323.
- Srere, P.A. (1959). The citrate cleavage enzyme. I. Distribution and purification. *J. Biol. Chem.* 234, 2544–2547.
- Srere, P.A., and Lipmann, F. (1953). An enzymatic reaction between citrate, adenosine triphosphate and Coenzyme A. *J. Am. Chem. Soc.* 75, 4874.
- Takahashi, H., McCaffery, J.M., Irizarry, R.A., and Boeke, J.D. (2006). Nucleo-cytosolic acetyl-coenzyme A synthetase is required for histone acetylation and global transcription. *Mol. Cell* 23, 207–217.
- Tollinger, C.D., Vreman, H.J., and Weiner, M.W. (1979). Measurement of acetate in human blood by gas chromatography: effects of sample preparation, feeding, and various diseases. *Clin. Chem.* 25, 1787–1790.
- Tsuchida, T., Takeuchi, H., Okazawa, H., Tsujikawa, T., and Fujibayashi, Y. (2008). Grading of brain glioma with 1-11C-acetate PET: comparison with 18F-FDG PET. *Nucl. Med. Biol.* 35, 171–176.
- Tu, B.P., Kudlicki, A., Rowicka, M., and McKnight, S.L. (2005). Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science* 310, 1152–1158.
- van den Berg, M.A., de Jong-Gubbels, P., Kortland, C.J., van Dijken, J.P., Pronk, J.T., and Steensma, H.Y. (1996). The two acetyl-coenzyme A synthetases of *Saccharomyces cerevisiae* differ with respect to kinetic properties and transcriptional regulation. *J. Biol. Chem.* 271, 28953–28959.
- Warburg, O. (1956a). On respiratory impairment in cancer cells. *Science* 124, 269–270.
- Warburg, O. (1956b). On the origin of cancer cells. *Science* 123, 309–314.
- Waterborg, J.H. (2002). Dynamics of histone acetylation in vivo. A function for acetylation turnover? *Biochem. Cell Biol.* 80, 363–378.
- Watkins, P.A., Maiguel, D., Jia, Z., and Pevsner, J. (2007). Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. *J. Lipid Res.* 48, 2736–2750.
- Wellen, K.E., Hatzivassiliou, G., Sachdeva, U.M., Bui, T.V., Cross, J.R., and Thompson, C.B. (2009). ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* 324, 1076–1080.

- Wise, D.R., Ward, P.S., Shay, J.E., Cross, J.R., Gruber, J.J., Sachdeva, U.M., Platt, J.M., DeMatteo, R.G., Simon, M.C., and Thompson, C.B. (2011). Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of α -ketoglutarate to citrate to support cell growth and viability. *Proc. Natl. Acad. Sci. USA* **108**, 19611–19616.
- Yao, Y.J., Ping, X.L., Zhang, H., Chen, F.F., Lee, P.K., Ahsan, H., Chen, C.J., Lee, P.H., Peacocke, M., Santella, R.M., and Tsou, H.C. (1999). PTEN/MMAC1 mutations in hepatocellular carcinomas. *Oncogene* **18**, 3181–3185.
- Yoshii, Y., Furukawa, T., Yoshii, H., Mori, T., Kiyono, Y., Waki, A., Kobayashi, M., Tsujikawa, T., Kudo, T., Okazawa, H., et al. (2009a). Cytosolic acetyl-CoA synthetase affected tumor cell survival under hypoxia: the possible function in tumor acetyl-CoA/acetate metabolism. *Cancer Sci.* **100**, 821–827.
- Yoshii, Y., Waki, A., Furukawa, T., Kiyono, Y., Mori, T., Yoshii, H., Kudo, T., Okazawa, H., Welch, M.J., and Fujibayashi, Y. (2009b). Tumor uptake of radiolabeled acetate reflects the expression of cytosolic acetyl-CoA synthetase: implications for the mechanism of acetate PET. *Nucl. Med. Biol.* **36**, 771–777.
- Yun, M., Bang, S.H., Kim, J.W., Park, J.Y., Kim, K.S., and Lee, J.D. (2009). The importance of acetyl coenzyme A synthetase for ^{11}C -acetate uptake and cell survival in hepatocellular carcinoma. *J. Nucl. Med.* **50**, 1222–1228.

Acetate Is a Bioenergetic Substrate for Human Glioblastoma and Brain Metastases

Tomoyuki Mashimo,^{1,2,3,11} Kumar Pichumani,^{4,11} Vamsidhara Vemireddy,^{1,2,3} Kimmo J. Hatanpaa,^{2,3,5} Dinesh Kumar Singh,^{2,3,6} Shyam Sirasanagandla,^{1,2,3} Suraj Nannepaga,^{2,3,6} Sara G. Piccirillo,^{3,6} Zoltan Kovacs,⁴ Chan Foong,⁵ Zhiguang Huang,⁷ Samuel Barnett,⁸ Bruce E. Mickey,^{2,3,8} Ralph J. DeBerardinis,^{2,9,10} Benjamin P. Tu,⁷ Elizabeth A. Maher,^{1,2,3,6,12,*} and Robert M. Bachoo^{1,2,3,6,12,*}

¹Department of Internal Medicine, UT Southwestern Medical Center, Dallas, TX 75390, USA

²Simmons Comprehensive Cancer Center, UT Southwestern Medical Center, Dallas, TX 75390, USA

³Annette G. Strauss Center for Neuro-Oncology, UT Southwestern Medical Center, Dallas, TX 75390, USA

⁴Advanced Imaging Research Center, UT Southwestern Medical Center, Dallas, TX 75390, USA

⁵Department of Pathology, UT Southwestern Medical Center, Dallas, TX 75390, USA

⁶Department of Neurology and Neurotherapeutics, UT Southwestern Medical Center, Dallas, TX 75390, USA

⁷Department of Biochemistry, UT Southwestern Medical Center, Dallas, TX 75390, USA

⁸Department of Neurological Surgery, UT Southwestern Medical Center, Dallas, TX 75390, USA

⁹McDermott Center for Human Growth and Development, UT Southwestern Medical Center, Dallas, TX 75390, USA

¹⁰Children's Medical Center Research Institute, UT Southwestern Medical Center, Dallas, TX 75390, USA

¹¹Co-first author

¹²Co-senior author

*Correspondence: elizabeth.maher@utsouthwestern.edu (E.A.M.), robert.bachoo@utsouthwestern.edu (R.M.B.)

<http://dx.doi.org/10.1016/j.cell.2014.11.025>

SUMMARY

Glioblastomas and brain metastases are highly proliferative brain tumors with short survival times. Previously, using ¹³C-NMR analysis of brain tumors resected from patients during infusion of ¹³C-glucose, we demonstrated that there is robust oxidation of glucose in the citric acid cycle, yet glucose contributes less than 50% of the carbons to the acetyl-CoA pool. Here, we show that primary and metastatic mouse orthotopic brain tumors have the capacity to oxidize [1,2-¹³C]acetate and can do so while simultaneously oxidizing [1,6-¹³C]glucose. The tumors do not oxidize [U-¹³C]glutamine. In vivo oxidation of [1,2-¹³C]acetate was validated in brain tumor patients and was correlated with expression of acetyl-CoA synthetase enzyme 2, ACSS2. Together, the data demonstrate a strikingly common metabolic phenotype in diverse brain tumors that includes the ability to oxidize acetate in the citric acid cycle. This adaptation may be important for meeting the high biosynthetic and bioenergetic demands of malignant growth.

INTRODUCTION

Malignant brain tumors are among the most intractable problems in cancer. Glioblastoma (GBM), the most common and aggressive primary tumor, has a median survival of 16 months. Despite intense clinical efforts at targeting various signaling pathways, putative driver mutations, and angiogenesis mecha-

nisms, no improvement in survival has emerged since 2005, with the addition of temozolomide to radiation as initial therapy (Cloughesy et al., 2014; Fine, 2014). Brain metastases, similarly, are aggressive tumors that affect ~200,000 patients per year in the United States (Lu-Emerson and Eichler, 2012) and usually occur late in the clinical course, often heralding end-stage disease. Treatment options are limited, and survival is measured in months (Owonikoko et al., 2014). Although GBM and brain metastases represent a broad range of cancer subtypes with distinct cellular origins and diverse genetic programs, they exhibit common metabolic characteristics that may be the result of reprogramming to enable rapid growth in the brain. Using ¹³C-NMR, we have previously shown in patients with GBM, lung, and breast cancer brain metastases that these tumors oxidize glucose in the citric acid cycle (CAC) to produce macromolecular precursors and energy (Maher et al., 2012). The metabolic complexity of these tumors is further reflected in the identification of a “bioenergetic substrate gap,” whereby a significant fraction of the acetyl-CoA pool is not derived from blood-borne glucose (Maher et al., 2012). The striking commonality of this finding among different grades of gliomas and metastatic tumors of diverse cellular origins prompted us to consider the possibility that an alternate or additional substrate(s) may serve as an important carbon source for generating CAC intermediates to support biosynthesis and bioenergetics in vivo.

Although the normal healthy adult brain relies almost exclusively on glucose as the major energy substrate, it can readily adapt to alternate fuels, including ketone bodies, short and medium chain fatty acids, and acetate (Ebert et al., 2003). Astrocytes are capable of supporting neuronal function by utilizing acetate as a metabolic substrate under conditions of limiting glucose supply, including diabetic hypoglycemia and chronic alcohol abuse (Cloughesy et al., 2014; Schurr, 2001;

Jiang et al., 2013). Because GBMs develop from the astroglial lineage, we hypothesized that these tumors retain the capacity to metabolize acetate during transformation. The most common brain metastases, in contrast, arise from organs that are not known to utilize substrates other than glucose. We speculated, that the unique brain microenvironment might drive tumors of diverse origins to utilize the same metabolic substrates to fuel aggressive growth. To test this hypothesis in vivo, we used human orthotopic tumor (HOT) mouse models of GBM and brain metastases and applied methods in intermediary metabolism for studying multiple substrates using ^{13}C -labeled nutrients (Malloy et al., 1988; Sherry et al., 1992). Coinfusion of ^{13}C -acetate and ^{13}C -glucose has been used extensively to study normal rodent brain metabolism in which differential handling of acetate and glucose by the glial and neuronal compartments can be demonstrated by ^{13}C -NMR of resected brain tissue. These methods enable direct tracing of the metabolic fate of infused substrates beyond simple uptake in the cell and therefore can be used to determine directly whether acetate can be oxidized by GBM and/or brain metastases in an orthotopic model in vivo.

Here, we report that mice harboring human GBM or brain metastases can completely oxidize acetate in the tumors. We have validated this finding in patient tumors by infusing ^{13}C -acetate in patients with GBM, breast cancer, and non-small cell lung cancer during surgical resection of their tumors and show that there is robust labeling of CAC intermediates by blood-borne ^{13}C -acetate.

In the article by Comerford et al. (2014) in this issue of *Cell*, the authors demonstrate a critical role for the nucleocytosolic acetyl-CoA synthetase enzyme, ACSS2, in hepatocellular carcinoma and broad immunoreactivity for ACSS2 in diverse human tumor types, including gliomas, breast cancer, and lung cancer. Here, we show that ACSS2 is upregulated in the HOT and primary human tumors, as well as a murine glioma model. In ACSS2 knockout mouse embryo fibroblasts (MEFs), ^{13}C -acetate fails to label CAC intermediates, and in human GBM neurospheres, stable ACSS2 knockdown leads to failure of self-renewal. These studies provide a potential mechanistic link between ACSS2 activity and in vivo acetate oxidation in tumors.

RESULTS

Glioblastomas Oxidize Acetate in the Citric Acid Cycle

The human orthotopic tumor (HOT) lines of GBM and brain metastases used in this study were each derived from an individual patient tumor, and implanted into the basal ganglia of NOD-SCID mice within 3 hr of surgical resection. All mouse experiments were approved by the Animal Resource Center, University of Texas Southwestern Medical Center (UTSW). Clinically symptomatic tumors arose within 2–4 months. The tumors, which are linked to UTSW Institutional Research Board (IRB)-approved collection of clinical information, were serially passaged and expanded in the mouse brain without adaptation to cell culture. Brain-only passaging helps ensure the preservation of the phenotypic, molecular and metabolic profiles of the human tumors and the tumor-stromal interactions, to the extent possible in an experimental system. We selected six HOT lines (UT-GBM1-6; Table S1 available online) that are representative of the most common

GBM molecular profiles (Brennan et al., 2013). Each line was generated at the time of the patient's initial diagnosis prior to any treatment and was studied here in early in vivo passage. A seventh HOT line (UT-GBM7), generated at the time of repeat surgery for tumor recurrence 15 months after initial resection in the same patient from which the UT-GBM6 HOT line was derived, was chosen to compare substrate utilization in the setting of recurrence and multimodality resistance. We have validated in vivo that UT-GBM6 is temozolomide (TMZ) sensitive, whereas UT-GBM7 is TMZ resistant (Sagiyama et al., 2014).

Representative histological sections from UT-GBM1 (Figure 1A) show an expansive mass (T) comprised of densely packed tumor cells and infiltration into the brain at the leading edges. In each mouse, the contralateral hemisphere served as a matched control for substrate utilization. It is referred to as “non-tumor bearing brain” (NT; Figure 1A) rather than “normal brain” because the brain surrounding a large tumor is subjected to mass effect, an increase in reactive astrocytes and diffusible factors from the tumor and/or blood, conditions which could potentially impact brain metabolism. ^{13}C -NMR analysis of NT brain provides a valuable internal control for each mouse because the NT and T are exposed to the same circulating concentrations of ^{13}C -glucose and ^{13}C -acetate and systemic condition of the mouse. Thus, differences in labeling patterns and substrate utilization between NT and T reflect tumor-specific handling of the substrates.

[1,6- ^{13}C]glucose and [1,2- ^{13}C]acetate were chosen for coinfusion because oxidation of each substrate produces distinct labeling patterns in CAC intermediates, enabling a direct comparison of substrate utilization in a given tissue (Malloy et al., 1988; Taylor et al., 1996; Deelchand et al., 2009). In the schema (Figure 1B), metabolism of [1,6- ^{13}C]glucose (blue circles) leads to labeling of carbon 3 in pyruvate, followed by production of acetyl-CoA labeled in position 2, which condenses with oxaloacetate (OAA), leading to labeling of carbon 4 in α -ketoglutarate (α -KG), glutamate, and glutamine during the first turn of the CAC, and in carbons 3 and 4 with subsequent turns. This labeling generates a singlet (S) and doublet 3, 4 (D34) in glutamate C4 (GLU4) and glutamine C4 (GLN4) (for example, Figure 1C). In contrast, metabolism of [1,2- ^{13}C]acetate leads to labeling of both carbons (red circles) of acetyl-CoA with subsequent labeling of carbons 4 and 5 of α -KG and glutamate in the first turn of the CAC, generating a doublet 4,5 (D45) (for example, Figure 1C). In subsequent turns of the cycle, labeling in carbons 3, 4, and 5 generates a doublet of doublets (quartet, Q) (for example, Figure 2A). Thus, the ^{13}C -NMR multiplet pattern in carbon 4 of glutamate (GLU4) reflects differential labeling of the acetyl-CoA pool and provides a direct and unequivocal readout of substrate metabolism, whereby D45 and Q report acetate oxidation and S and D34 report glucose oxidation (Sherry et al., 1992). The fractional amount of each multiplet in GLU4 can be obtained by determining the area of each multiplet relative to the total spectral area of GLU4, in which the areas of S, D34, D45, and Q sum to one (Marin-Valencia et al., 2012b).

The NT brain ^{13}C -NMR spectrum (Figure 1C) shows high signal to noise and well-resolved multiplets arising from ^{13}C - ^{13}C coupling. Oxidation of both glucose and acetate in NT brain is demonstrated by the presence of S and D34 from glucose and

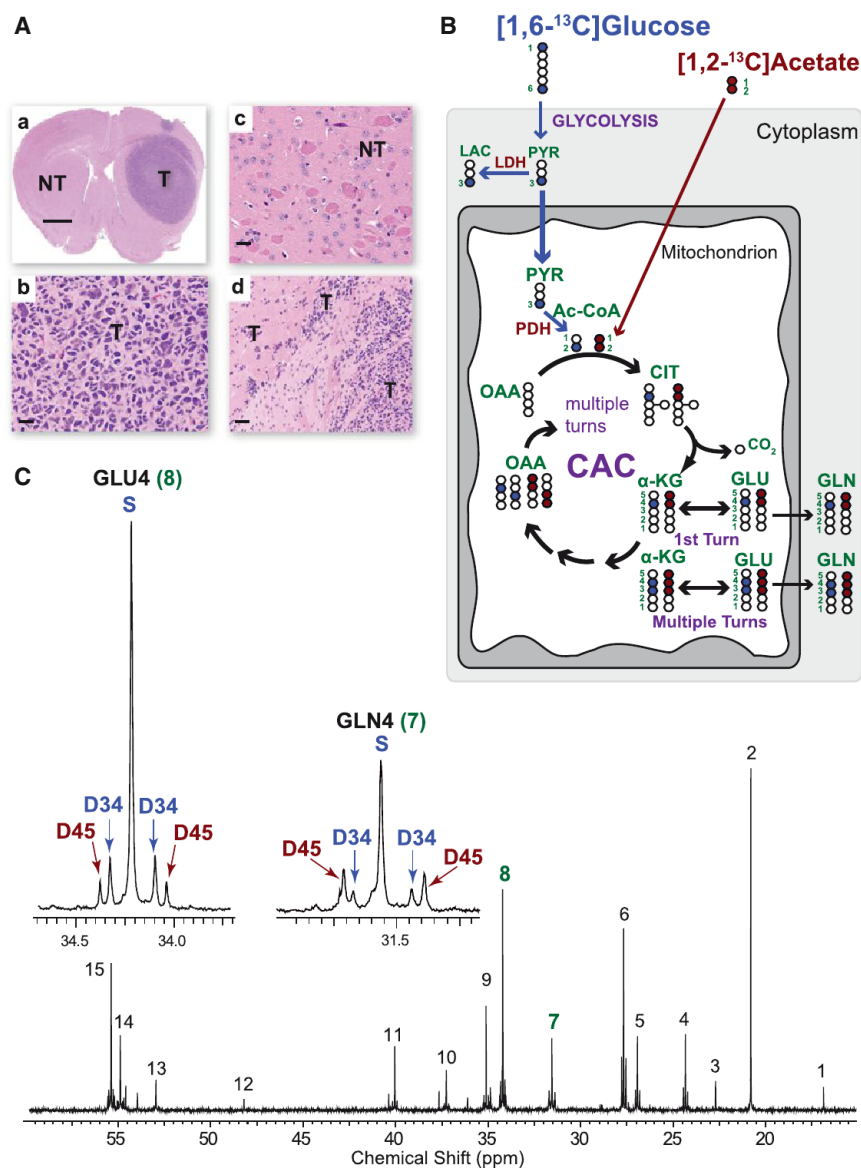


Figure 1. Metabolism of Coinfused [1,6- 13 C]Glucose and [1,2- 13 C]Acetate

(A) Low-power hematoxylin & eosin (H&E) images of a GBM H0T mouse brain at the time of co-infusion. (a) A large tumor (T) mass is seen in the right hemisphere. The left hemisphere is designated nontumor (NT) brain. Scale bar, 3 mm. High-power images of (b) tumor (T), (c) NT brain (scale bars, 10 μ m), and (d) tumor that infiltrates brain at the edge of the mass (scale bar, 20 μ m).

(B) Schema showing the fate of individual carbons from infused [1,6- 13 C]glucose (blue-filled circles, 13 C) and [1,2- 13 C]acetate (red filled circles, 13 C) through the first turn of the CAC and labeling in α -KG, GLU, and GLN after multiple turns. Open circles, 12 C. Numbers refer to carbon positions. Abbreviations: LAC, lactate; Ac-CoA, acetyl CoA; CIT, citrate; α -KG, α -ketoglutarate; GLU, glutamate; GLN, glutamine; OAA, oxaloacetate; PDH, pyruvate dehydrogenase; PYR, pyruvate.

(C) 13 C-NMR spectrum from NT brain after co-infusion. Insets are GLU4 and GLN4. Singlet (S) and doublet 3,4 (D34) in blue are generated from 13 C-glucose metabolism, and doublets 4,5 (D45) in red are generated from 13 C-acetate metabolism. The color scheme is the same in all figures. Chemical shift assignments are the same in Figures 2 and 3: 1, alanine C3; 2, lactate C3; 3, N-acetylaspartate (NAA) C6; 4, gamma-aminobutyric acid (GABA) C3; 5, glutamine C3; 6, glutamate C3; 7, glutamine C4; 8, glutamate C4; 9, GABA C2; 10, aspartate C3; 11, GABA C4; 12, taurine (?); 13, aspartate C2; 14, glutamine C2; 15, glutamate C2.

See also Table S1.

D45 from acetate in GLU4 (Figure 1C, inset). Although the same multiplets are present in glutamine C4 (GLN4), the pattern of D45 and D34 is different from that in GLU4, recapitulating the well-recognized pattern that results from the differential handling of acetate and glucose in glia and neurons (Taylor et al., 1996). The 13 C-NMR tumor spectrum from the same mouse (Figure 2A, inset) has several notable differences. First, there is a marked increase in D45 and the presence of Qs, indicating increased oxidation of acetate in the tumor when compared to the NT brain. Second, the similarity between the GLU4 and GLN4 labeling pattern (prominent D45 and small or absent D34) provides evidence that glutamine is being derived from glutamate in the tumor.

To determine the biological variability of acetate utilization within individual mice with tumors derived from the same parental line (UT-GBM1), we coinfused an additional four mice. The resulting 13 C-NMR spectral patterns of GLU4 and GLN4 in NT brain and tumor were almost indistinguishable from the NT

brain and tumor spectra presented in Figures 2A and 2B (data not shown). To quantify the relative contribution of each substrate to the labeling in GLU4, we calculated the acetate-to-glucose ratio, (D45+Q) to (S+D34). This has been used extensively in brain metabolism, where the differential labeling of multiplets enables discrimination of glial and neuronal metabolism, because astrocytes, but not neurons, are capable of oxidizing acetate (Deelchand et al., 2009; Marin-Valencia et al., 2012b). In applying this analysis to brain tumors, we assume a single compartment comprised of tumor cells, based on histological analysis. Thus, the acetate-to-glucose ratio in tumor GLU4 reflects the relative contribution of each labeled substrate. The ratio was calculated for the five mice, NT brain and tumor (Figure 2B).

In NT brain, glucose is the major substrate being oxidized with an acetate-to-glucose ratio of $15.8\% \pm 1.2\%$ to $84.2\% \pm 1.1\%$ (ratio 0.19 ± 0.02), achieved with 13 C-glucose fractional enrichment of $32\% \pm 5\%$ in the blood and plasma acetate levels that rose approximately 4-fold (from 0.177 ± 0.02 mM to 0.619 ± 0.13 mM) during infusion. The acetate-to-glucose ratio is remarkably similar to that previously reported in normal adult mouse brain (calculated as 14% to 86%, ratio 0.16; Marin-Valencia

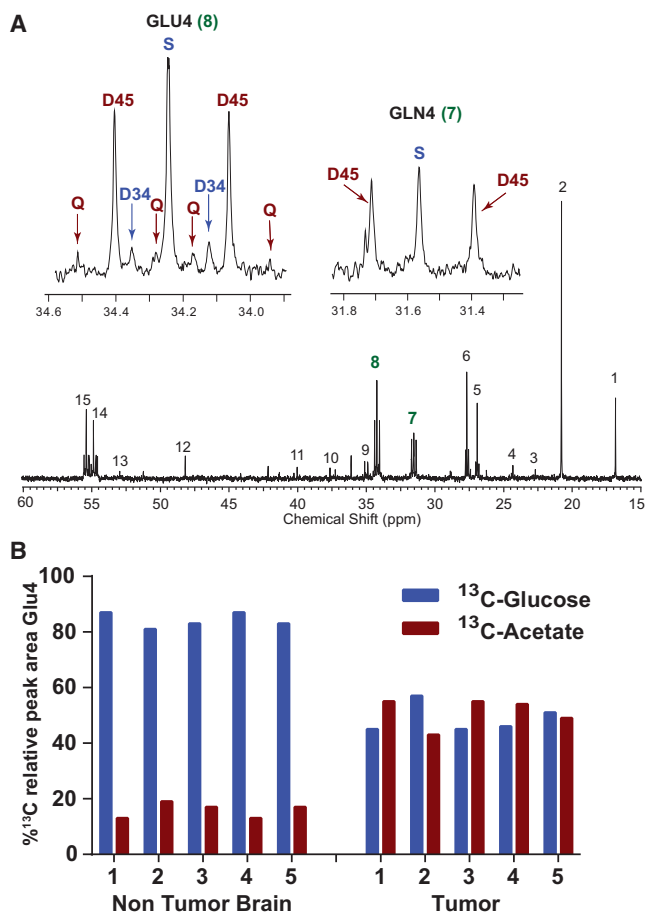


Figure 2. Metabolism of Coinfused [1,6-¹³C]Glucose and [1,2-¹³C]Acetate in GBM versus NT Brain

(A) ¹³C-NMR GBM tumor spectrum after coinfusion. Inserts are GLU4 and GLN4. Chemical shift assignments are the same as in Figure 1.

(B) Relative percent labeling of GLU4 by acetate and glucose in five replicates of NT brain and T from UT-GBM1. Peak areas in GLU4 were measured from the ¹³C-NMR spectra. The contribution of ¹³C-glucose (S+D34) (blue bars) and ¹³C-acetate (D45+Q) (red bars) is expressed as a percent of the total peak area in GLU4.

See also Figure S1 and Table S1.

et al., 2012b). The acetate-to-glucose ratio in the five tumors was $51.8\% \pm 2.9\%$ to $48.3\% \pm 2.9\%$ (ratio 1.02 ± 0.09), which was 5-fold higher than in NT brain ($p < 0.001$), demonstrating a significant shift toward acetate oxidation in the tumor.

Direct analysis of the multiplet ratios in GLU4 does not provide information about the contribution of unlabeled substrates in the acetyl-CoA pool, which could be playing an important role in fueling the tumor. To determine the size of the unlabeled pool, we applied a validated non-steady-state algorithm to the spectral analysis, which takes into account the total multiplet area in the glutamate C3 and C4 and the levels of infused substrates in the blood (Malloy et al., 1988, 1990b). The fraction of acetyl-CoA due to acetate increased from $6.0\% \pm 1.2\%$ in NT brain to $24.7\% \pm 2\%$ in the tumor ($p < 0.0001$). This was associated with a decrease in the fractional contribution from glucose

from $75.8\% \pm 3.7\%$ in NT brain to $50.4\% \pm 9.1\%$ in the tumor ($p < 0.002$), whereas the unlabeled fraction was not significantly different between NT brain ($18.2\% \pm 4.0\%$) and tumor ($25.5\% \pm 8.0\%$, $p > 0.05$). The decrease in glucose oxidation between NT brain and tumor may, at least in part, reflect a loss of the neuronal contribution to glucose utilization.

Having established the reproducibility of biological replicates in UT-GBM1 that has *PDGFR* overexpression/*PTEN* deletion, we next addressed the question of whether the ability to oxidize acetate is a function of the dominant oncogenes and/or tumor suppressor genes in the tumor. Mice from five additional GBM HOT lines (UT-GBM2-6; Table S1) were coinfused with [1,2-¹³C]acetate and [1,6-¹³C]glucose. The acetate-to-glucose ratios were calculated for these lines. No differences were found among the NT brains, and in each line, tumor acetate oxidation was higher than in NT brain. The relative ¹³C-acetate contribution to GLU4 ranged from 18.8% to 52% ($32\% \pm 5\%$) in the tumor. Differences in the percent acetate oxidation among the individual lines could not be attributed to amplification of *EGFR* or *PDGFRα*, presence of the *EGFR* VIII mutant, loss of *PTEN*, or loss of *INK4a/ARF* (data not shown).

Because all six GBM HOT lines were generated from patient tumors prior to the initiation of treatment, we next studied the tumor line generated from recurrent tumor (UT-GBM7) and compared it directly with the initial GBM HOT line from the same patient (UT-GBM6). Nearly identical labeling patterns in GLU4 and GLN4 were identified (Figure S1), with dominance of the D45, an acetate-to-glucose ratio 25% to 75% in the initial tumor (UT-GBM6), and 27% to 73% in the recurrent tumor (UT-GBM7). The relative avidity for acetate and glucose is remarkably similar in these tumors that were derived independently more than a year apart and after extensive treatment and acquisition of chemo- and radioresistance.

Brain Metastases from Diverse Cell Types Oxidize Acetate

To determine whether acetate oxidation is a feature of nonglial brain tumors, we studied five brain metastasis HOT tumor lines: breast cancer (estrogen receptor [ER] and progesterone receptor [PR] negative and HER2 positive), non-small cell lung cancer (no mutations in *EGFR*, *ALK*, or *KRAS*), clear cell renal cell carcinoma (*VHL*—/—), melanoma (*BRAF*^{V600E} mutated), and endometrial cancer, a tumor which rarely metastasizes to the brain. There is a remarkable preservation of the signature histopathological and molecular features of the human brain metastases in the HOT models (see examples in Figure 3). Of note, the breast and lung cancer HOT lines were generated from the human tumors infused with ¹³C-glucose during surgery that we reported previously (Maher et al., 2012), each of which oxidized glucose but also showed a significant bioenergetic substrate gap. The study of these tumors here as orthotopic mouse models provided a unique opportunity to assess metabolic phenotype fidelity in these models.

The ¹³C-NMR spectral patterns from the five individual tumor lines and matching NT brains following coinfusion of [1,6-¹³C] glucose and [1,2-¹³C]acetate recapitulated the findings in GBM with remarkable similarity. Representative GLU4 and GLN4 labeling from the breast cancer (Figure 3A) and melanoma

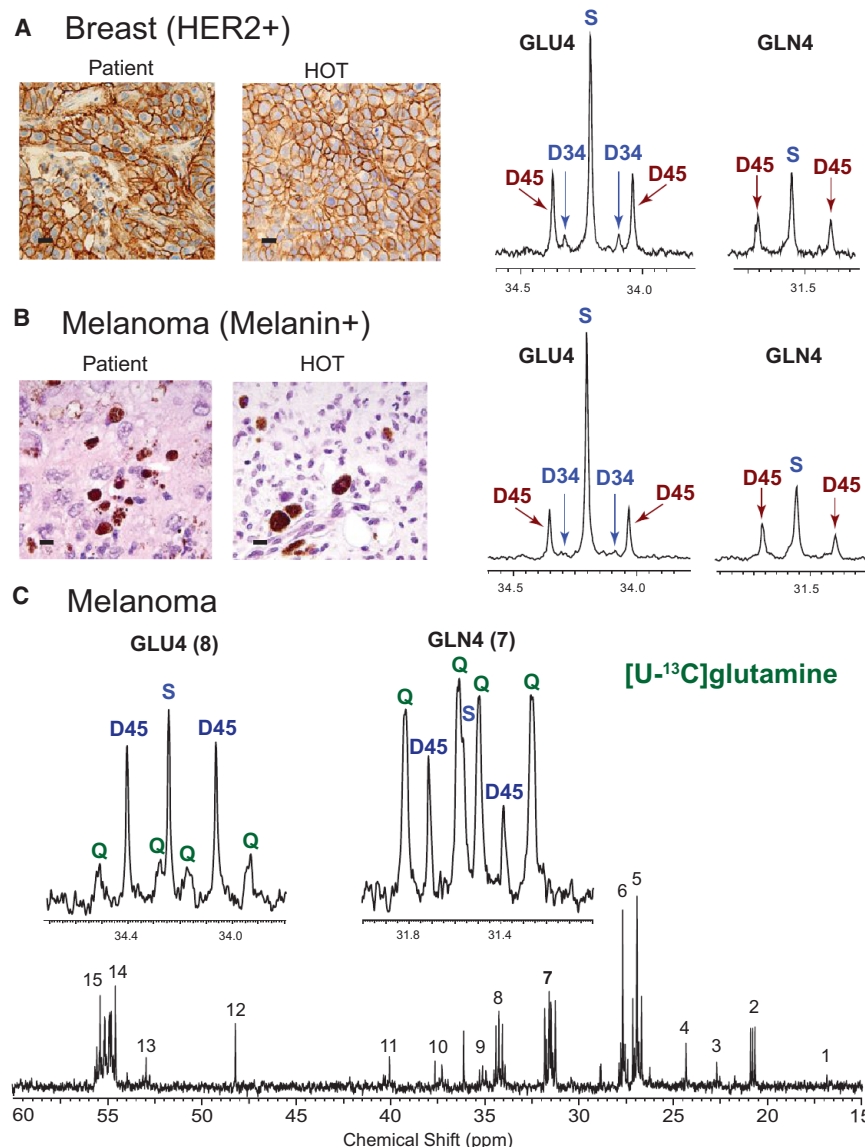


Figure 3. Oxidation of ¹³C-Acetate, but Not ¹³C-Glutamine, in Brain Metastases

(A and B) Coinfusion of [1,6-¹³C]glucose and [1,2-¹³C]acetate in HOTA models from breast cancer brain metastasis (ER-, PR-, HER2+) (A) and melanoma (*BRAF*^{V600E} mutant) (B). Scale bars, 10 μ m. Patient tumor (left) is compared with the HOTA mouse tumor generated from the same patient tumor (right) for HER2 in breast (A) and melanin in melanoma (B). The GLU4 and GLN4 ¹³C-NMR profiles from the HOTA tumors show similar labeling patterns with prominent D45 generated from ¹³C-acetate oxidation. The presence of S and D34 indicates that ¹³C-glucose was oxidized simultaneously. Note the similar pattern in GLN4 (prominent D45) in both tumor spectra. (C) Infusion of [U-¹³C]glutamine in the melanoma HOTA model. Insets are GLU4 and GLN4. Note the different labeling pattern in GLU4 and GLN4. Prominent labeling was not detected in aspartate (C3 in position 10 and C2 in position 13 in the full spectrum) or malate (MAL) (not labeled). See also Figure S2 and Table S1.

and three brain metastasis HOTA lines (non-small cell lung cancer, melanoma, and endometrial cancer) to determine whether glutamine could be directly oxidized in these tumors. Nearly identical spectra were obtained following infusion of [U-¹³C]glutamine in the three tumor lines (Figures 3C, S2A, and S2B) and in GBM (data not shown). GLN4 was dominated by large Qs, which is consistent with the presence of ¹³C-glutamine in the tumor. The presence of smaller Qs in GLU4 suggests that there is some direct exchange of ¹³C-glutamine with glutamate. However, the GLU4 in each tumor was dominated by D45, a pattern that cannot be produced simply from exchange with glutamine. Moreover, prominent Qs in malate and aspartate were not visible in the spectra (Figures 3C and S2), as would have been expected if [U-¹³C]glutamine had exchanged with glutamate, which then exchanged with α -KG and subsequently was fully oxidized in the CAC. An alternative route for GLU4 ¹³C labeling in the tumor was more likely than direct glutamine oxidation. The multiplet patterns in lactate and alanine from liver, NT brain, and tumor from the lung brain metastasis HOTA line (Figure S2B) are consistent with production of ¹³C-glucose from infused [U-¹³C]glutamine outside the CNS, likely in the liver, with subsequent ¹³C-glucose oxidation in NT brain and tumor. This is supported by the presence of ¹³C-labeled glucose in the liver ¹³C-NMR spectrum (Figure S2C). Taken together, the labeling patterns in the HOTA GBM and brain metastasis tumors following infusion of [U-¹³C]glutamine further extends our previous finding in GBM that glutamine is taken up by the tumors but is not directly oxidized in the CAC in vivo.

(Figure 3B) reveal prominent D45 consistent with oxidation of acetate, as well as a large singlet (S) and small D34, consistent with glucose oxidation. The acetate-to-glucose ratio in GLU4 was calculated for NT brain and tumor. Similar to the GBMs, there was a significant increase in the fraction of acetate being oxidized in the tumors, ranging from 21 to 42% (29% \pm 3%). Calculation of the fractional contribution to the acetyl-CoA pool for these tumors yielded 14.5% \pm 3.9% from ¹³C-acetate, 49.2% \pm 5.6% from ¹³C-glucose and 37% \pm 4.3% from the unlabeled substrate.

Glutamine Is Not Directly Oxidized in the CAC of GBM and Brain Metastases In Vivo

We have previously shown that GBM HOTA tumors do not oxidize glutamine in the CAC in vivo (Marin-Valencia et al., 2012c). Here, we examined an additional two GBM HOTA lines to increase the molecular diversity of the lines being examined

and three brain metastasis HOTA lines (non-small cell lung cancer, melanoma, and endometrial cancer) to determine whether glutamine could be directly oxidized in these tumors. Nearly identical spectra were obtained following infusion of [U-¹³C]glutamine in the three tumor lines (Figures 3C, S2A, and S2B) and in GBM (data not shown). GLN4 was dominated by large Qs, which is consistent with the presence of ¹³C-glutamine in the tumor. The presence of smaller Qs in GLU4 suggests that there is some direct exchange of ¹³C-glutamine with glutamate. However, the GLU4 in each tumor was dominated by D45, a pattern that cannot be produced simply from exchange with glutamine. Moreover, prominent Qs in malate and aspartate were not visible in the spectra (Figures 3C and S2), as would have been expected if [U-¹³C]glutamine had exchanged with glutamate, which then exchanged with α -KG and subsequently was fully oxidized in the CAC. An alternative route for GLU4 ¹³C labeling in the tumor was more likely than direct glutamine oxidation. The multiplet patterns in lactate and alanine from liver, NT brain, and tumor from the lung brain metastasis HOTA line (Figure S2B) are consistent with production of ¹³C-glucose from infused [U-¹³C]glutamine outside the CNS, likely in the liver, with subsequent ¹³C-glucose oxidation in NT brain and tumor. This is supported by the presence of ¹³C-labeled glucose in the liver ¹³C-NMR spectrum (Figure S2C). Taken together, the labeling patterns in the HOTA GBM and brain metastasis tumors following infusion of [U-¹³C]glutamine further extends our previous finding in GBM that glutamine is taken up by the tumors but is not directly oxidized in the CAC in vivo.

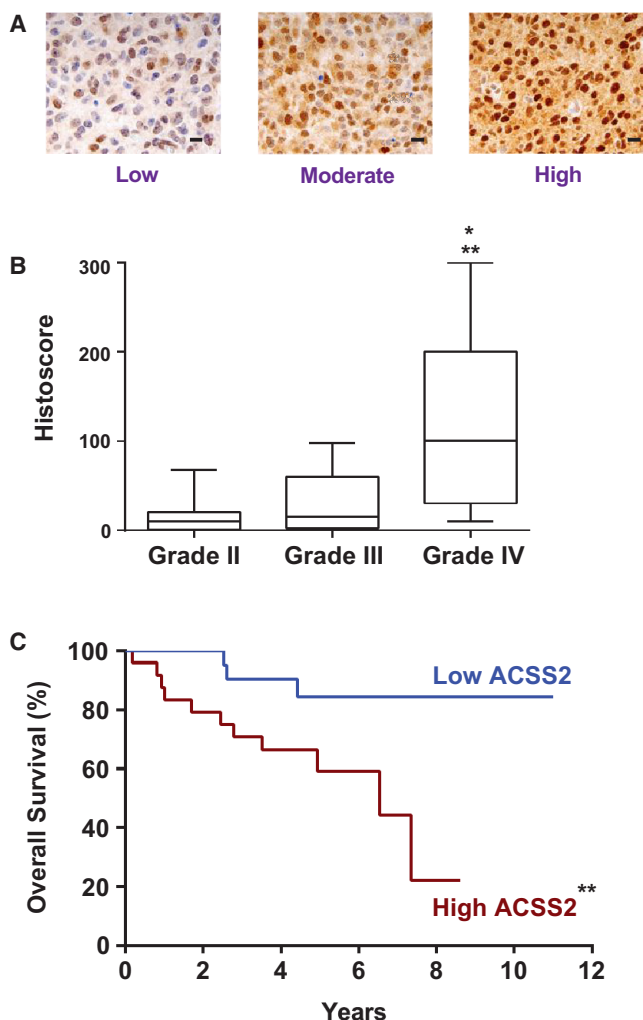


Figure 4. Immunoreactivity to ACSS2 Is Correlated with Glioma Grade and Survival of Grade II and III Gliomas

(A) Representative sections from glioma TMA showing the range of ACSS2 staining. Low, fewer than 50% of tumor cells are positive, and intensity of staining is one (scale 0–3); moderate, 75% positive, intensity 1–3; high, 100% positive and intensity 2–3.

(B) Box and whisker plot of ACSS2 histoscore for WHO grades II, III, and IV gliomas. **grade II versus grade IV ($p < 0.001$), and *grade III versus grade IV ($p < 0.01$).

(C) Kaplan-Meier curve of grade II and III astrocytomas and oligoastrocytomas. High versus low ACSS2 staining ($n = 25$ each group) based on the median histoscore. ** $p < 0.001$.

See also Figure S3A.

ACSS2 Upregulation Is Associated with More Aggressive Disease in Glioma

ACSS2 is a critical enzyme for converting acetate to acetyl-CoA in murine models of liver cancer (Comerford et al., 2014) and was assessed here for a potential role in acetate metabolism in brain tumors. ACSS2 immunohistochemistry (IHC) showed moderate to high expression in 6 of 7 GBM and all brain metastasis HOT lines (Figure S3). It was notable that the GBM with the lowest

fractional acetate oxidation (~18%) had the lowest ACSS2 expression by IHC (histoscore = 50). ACSS2 expression in our clinically annotated glioma tissue microarray (TMA) was variable (Figure 4A) but significantly higher in GBM than the grade II and III gliomas (Figure 4B). Among the grade II and III gliomas, shorter survival time was associated with higher ACSS2 staining (Figure 4C). ACSS1, ACSS3, sterol regulatory element binding protein (SREBP-1) expression in the TMA was not correlated with ACSS2 expression or with survival in any of the clinical subgroups (data not shown).

To determine whether ACSS2 has a direct impact on GBM growth, primary neurosphere cultures derived from two independent GBM HOT lines were infected with retroviruses expressing a small hairpin RNAi (shRNAi)-targeting ACSS2 for knockdown (KD) or a scrambled shRNAi (SCR) for control. Comerford et al. (2014) have extensively characterized these retroviruses. Infection with ACSS2-KD shRNAi, but not SCR, resulted in cell death (Figure 5A). Surviving GBM cells failed to form neurospheres ($2\% \pm 0.5\%$ versus $23\% \pm 4\%$, $p < 0.001$) and thus could not be assessed for in vivo tumorigenicity.

ACSS2 Expression Is Associated with Transformation and Enables ^{13}C -Acetate Incorporation into Glutamate

Next, we investigated whether ACSS2 has a potential role in glioma tumorigenicity. Primary astrocyte cultures were generated from multiallele conditional mice carrying glioma-relevant mutations ($p53^{fl/fl}$, $PTEN^{fl/fl}$, and $LSL\text{-}BRAF^{V600E}$) that activate the PI3K/AKT and MAPK/ERK pathways. In this ex vivo model, successive accumulation of mutations was associated with increased ACSS2 expression, with the triple-mutant cells, $p53^{-/-}PTEN^{-/-}BRAF^{V600E}$, having the highest expression (Figure 5B). High-grade gliomas generated from intracranial implantation of the triple-mutant cells were able to oxidize acetate when coin-fused with ^{13}C -acetate and ^{13}C -glucose (Figure 5C). It is notable that, in contrast to the tumor suppressor-deficient, immortalized primary astrocytes, wild-type (WT) astrocytes markedly downregulated expression of all ACSS enzymes under standard culture conditions (Figure 5D), making it an unreliable experimental system for further in vitro acetate metabolic studies.

To determine whether there is a direct mechanistic link between ACSS2 and oxidation of ^{13}C -acetate, we investigated the extent of ^{13}C -acetate incorporation into glutamate in ACSS2 KO MEFs. The KO MEFs exhibited very little incorporation of ^{13}C -acetate into glutamate, whereas reintroduction of the WT ACSS2 gene into these cells significantly increased acetate incorporation into glutamate (Figure S3). This direct genetic gain-of-function experiment provides a compelling link between ACSS2 and acetate oxidation in the CAC.

^{13}C -Acetate Is Oxidized in Patient Brain Tumors

To validate in humans the finding that acetate can be oxidized in the orthotopic models, we next infused $[1,2\text{-}^{13}\text{C}]\text{acetate}$ in four patients (two GBM, one breast cancer brain metastasis, and one non-small cell lung cancer brain metastasis) during surgical resection of their tumors. Despite the diversity of tumor types, nearly identical ^{13}C -NMR spectra with excellent signal-to-noise ratio and robust labeling in glutamate and glutamine were obtained in the four tumors (Figures 6, 7, and S4). The dominance

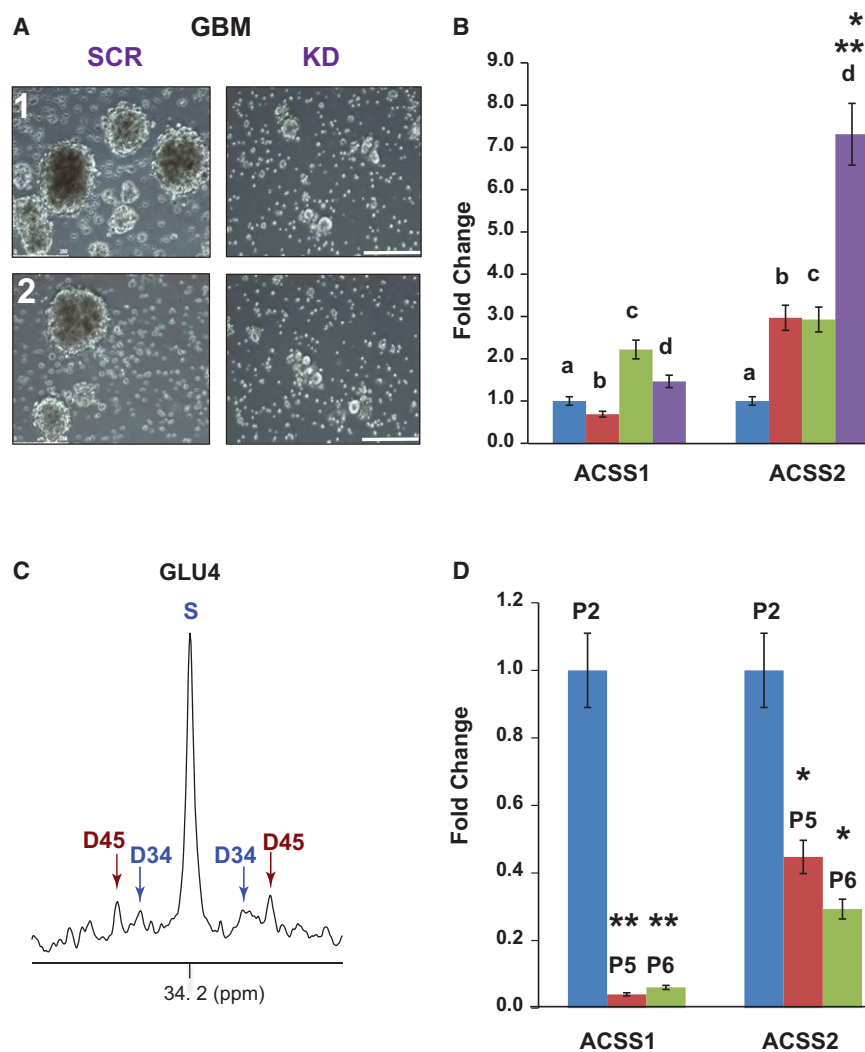


Figure 5. Expression of ACSS2 Is Linked to GBM Growth and Malignant Potential

(A) Comparison of primary GBM cultures infected with retroviruses expressing ACSS2 shRNAi (KD) or scrambled shRNAi (SCR). Two independent cultures are shown. Neurospheres are visible in the SCR, cultures but not in KD cultures. Scale bars, 250 μ m.

(B) Fold change (qRT-PCR) of ACSS1 and ACSS2 mRNA in primary conditional (floxed) astrocyte cultures after infection with adenovirus to produce (a) *p53*^{-/-}, (b) *p53*^{-/-}, *PTEN*^{-/-}, (c) *p53*^{+/+}, *PTEN*^{-/-}, *BRAF*^{V600E}, and (d) *p53*^{-/-}, *PTEN*^{-/-}, *BRAF*^{V600E}. **p* < 0.01 versus b or c; ***p* < 0.001 versus a.

(C) Glutamate C4 multiplets from ¹³C-NMR spectrum of an intracranial tumor arising from *p53*^{-/-}, *PTEN*^{-/-}, and *BRAF*^{V600E} astrocytes after co-infusion of ¹³C-acetate and ¹³C-glucose.

(D) Fold change (qRT-PCR) of ACSS1 and ACSS2 mRNA in primary astrocytes in culture at passages 2 (P2), 5 (P5), and 6 (P6). **p* < 0.01; ***p* < 0.005 versus P2.

See also Figure S3B.

DISCUSSION

Determining the carbon source(s) for biosynthesis and bioenergetics in rapidly proliferating tumors in vivo represents a critical step toward identifying potential vulnerabilities for therapeutic targeting. Our initial ¹³C-NMR data from ¹³C-glucose infusion in brain tumor patients (Maher et al., 2012) demonstrated a complex metabolic tumor phenotype that had not previously been reported; excess production of lactate (the Warburg effect) occurring

of the D45 and presence of quartets (Q) in GLU4 and GLN4 is direct evidence that acetate was oxidized in the CAC. Moreover, the same multiplet pattern in glutamine demonstrates that this metabolite was generated from glutamate that had originated as blood-borne acetate. ¹³C-acetate enrichment in the blood was 88.5% \pm 6.4%. Less than 3% (2.4% \pm 1.1%) ¹³C-glucose was recovered in the blood, demonstrating that the labeling in the tumors was not due to ¹³C-glucose production outside the CNS. The fractional contribution of [1,2-¹³C]acetate to the acetyl-CoA pool in the four tumors was 47.8% \pm 3.8%. Unlabeled substrate accounted for 51.3% \pm 2.6% of the acetyl-CoA pool, with 1.7% \pm 0.4% coming from [2-¹³C]acetate, consistent with natural abundance ¹³C. The unlabeled fraction is likely due in large part to circulating glucose, based on data from our previous study of brain tumor patients infused with ¹³C-glucose (Maher et al., 2012) and the ability of the orthotopic mouse tumors to co-oxidize glucose and acetate. ACSS2 immunoreactivity ranged from moderately to strongly positive in the four tumors with specificity demonstrated in the lung brain metastasis (Figure 7), where tumor, but not surrounding stroma, is labeled.

simultaneously with CAC oxidation of glucose as well as oxidation of another unidentified substrate(s), termed the "bioenergetic substrate gap." Here, we show that brain tumors from widely diverse cellular origins have the capacity to oxidize infused ¹³C-acetate. This finding in GBMs is perhaps not surprising given that glial cells, the presumptive GBM cells of origin, are well known to be capable of oxidizing acetate. This suggests that, despite the extensive molecular reprogramming that is a hallmark of GBM, the ability to oxidize acetate is preserved, if not enhanced, in the tumors. ¹¹C-acetate positron emission tomography (PET) positivity has been reported in gliomas (Yamamoto et al., 2008), consistent with our finding that acetate is actively taken up by the proliferating tumor cells. Nutrient uptake, however, although being of significant clinical utility as an imaging biomarker, provides no information about entry of the nutrient into specific metabolic pathways. A major advantage of ¹³C-NMR, as demonstrated here, is that there are multiple internal controls within a single spectrum for examination of the relative activity of specific metabolic pathways. For example, the pyruvate pool is read out

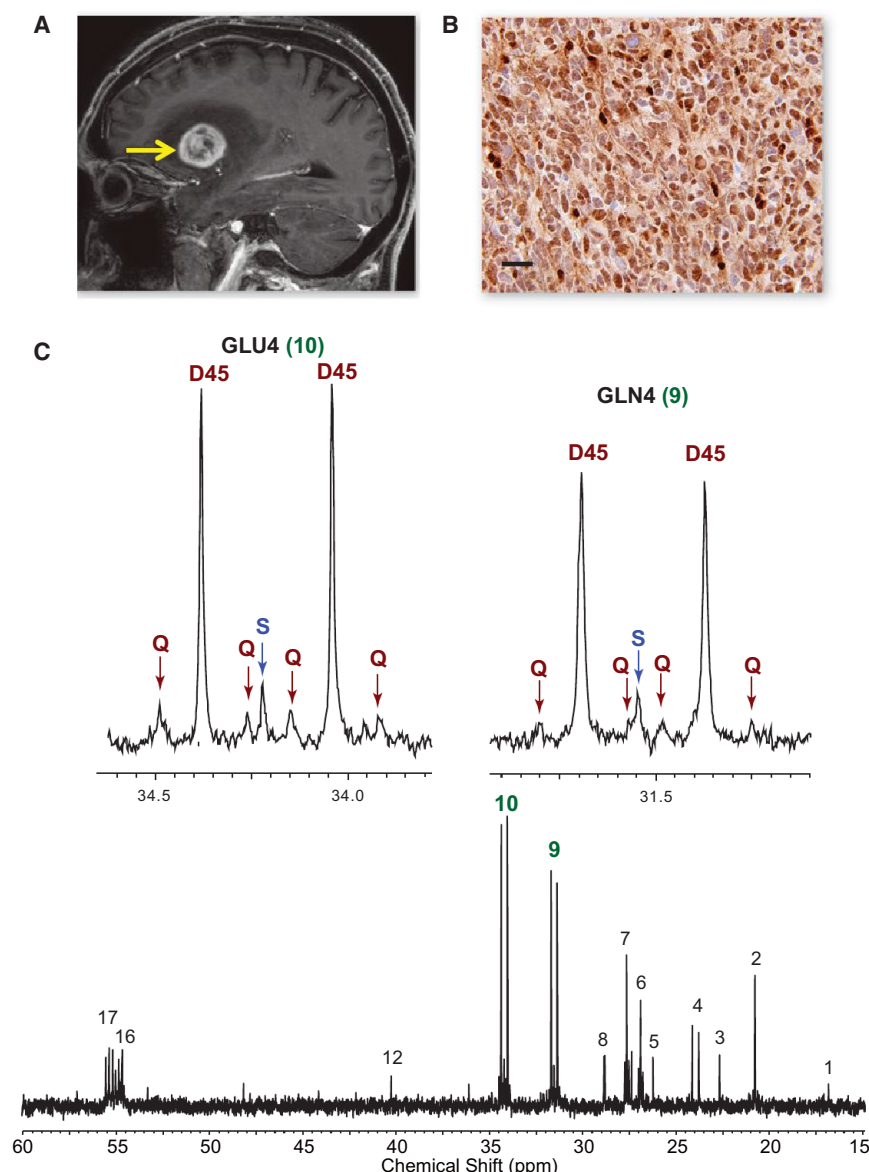


Figure 6. Oxidation of [1,2- ^{13}C]-Acetate in a Patient with GBM

(A) Preoperative sagittal image from gadolinium-enhanced MRI shows a large enhancing tumor in the left frontotemporal region (arrow).

(B) Strong ACSS2 immunoreactivity in the tumor. Scale bar, 10 μm .

(C) ^{13}C -NMR spectrum with GLU4 and GLN4 insets. Note the prominent D45 and Qs reflecting robust ^{13}C -acetate oxidation. Abbreviations are the same as in Figure 2. Chemical shift assignments are the same in Figure 7: 1, alanine C3; 2, lactate C3; 3, NAA C6; 4, acetate C2; 5, unassigned; 6, glutamine C3; 7, glutamate C3; 8, unassigned; 9, glutamine C4; 10, glutamate C4; 12, NAA C3; 16, glutamine C2; 17, glutamate C2.

in rapidly proliferating tumors. It would ensure the availability of an adequate pool of carbons to generate CAC intermediates and support the high bioenergetic demands of growth while a large fraction of the glucose is being diverted to lactate. In this study there was no evidence of direct ^{13}C -glutamine oxidation in the orthotopic models, which may reflect differential substrate handling in brain tumors.

Normal plasma acetate levels in non-fasted humans range between 0.05 mM (Tollinger et al., 1979) and 0.18 mM (Skutches et al., 1979). Under normal resting conditions, circulating acetate levels may contribute up to 10%–15% of the basal energy demands of brain astrocytes (Dienel and Cruz, 2006). Circulating free acetate is generated by fermentation of carbohydrates in the gut by commensal anaerobes and by the liver under ketogenic conditions (low glucose) as a final product of fatty acid oxidation or by ethanol metabolism in heavy drinkers (Jiang et al., 2013). It has recently been reported that 3 hr after oral administration of ^{13}C -inulin to mice, doublets of C4 glutamate and glutamine can be clearly resolved in brain by high-resolution magic angle spinning MR spectroscopy from brain uptake of circulating ^{13}C -labeled acetate (Frost et al., 2014). Moreover, the relative contribution of acetate to brain metabolism has been shown to increase following brain injury, which limits glucose oxidation (Bartnik-Olson et al., 2010). This finding may be analogous to brain tumor regions with a paucity of active neurons and limited PDH activity as a result of intratumoral hypoxia. Taken together, there is clear evidence that normal circulating acetate levels may be adequate to contribute to the metabolic demands of the tumor.

In order for the tumor cell to metabolize acetate, the cell must upregulate ACSS2, the enzyme critical for converting acetate to acetyl-CoA, as shown in hepatocellular carcinoma in the accompanying manuscript (Comerford et al., 2014). The increased

by labeling in lactate and alanine, whereas the acetyl-CoA pool is read out by labeling in glutamate.

The finding that brain metastases from a wide spectrum of cellular origins also oxidize acetate, however, was not anticipated because organs that most frequently give rise to tumors that metastasize to the brain (lung, breast, kidney, and melanoma) are not known to show significant ^{11}C -acetate uptake on PET. The data suggest that the ability to oxidize acetate is either a unique adaptation to the brain microenvironment or a more general property of tumor cells. In support of the latter possibility, several non-CNS tumors, including hepatocellular and prostate cancer, show avid differential ^{11}C -acetate uptake on PET relative to the normal organ (Ho et al., 2003; Oyama et al., 2002).

“Co-oxidation” of acetate and glucose was seen in all of the GBM and brain metastasis HOT models, raising the possibility that substrate co-oxidation might be an adaptive mechanism

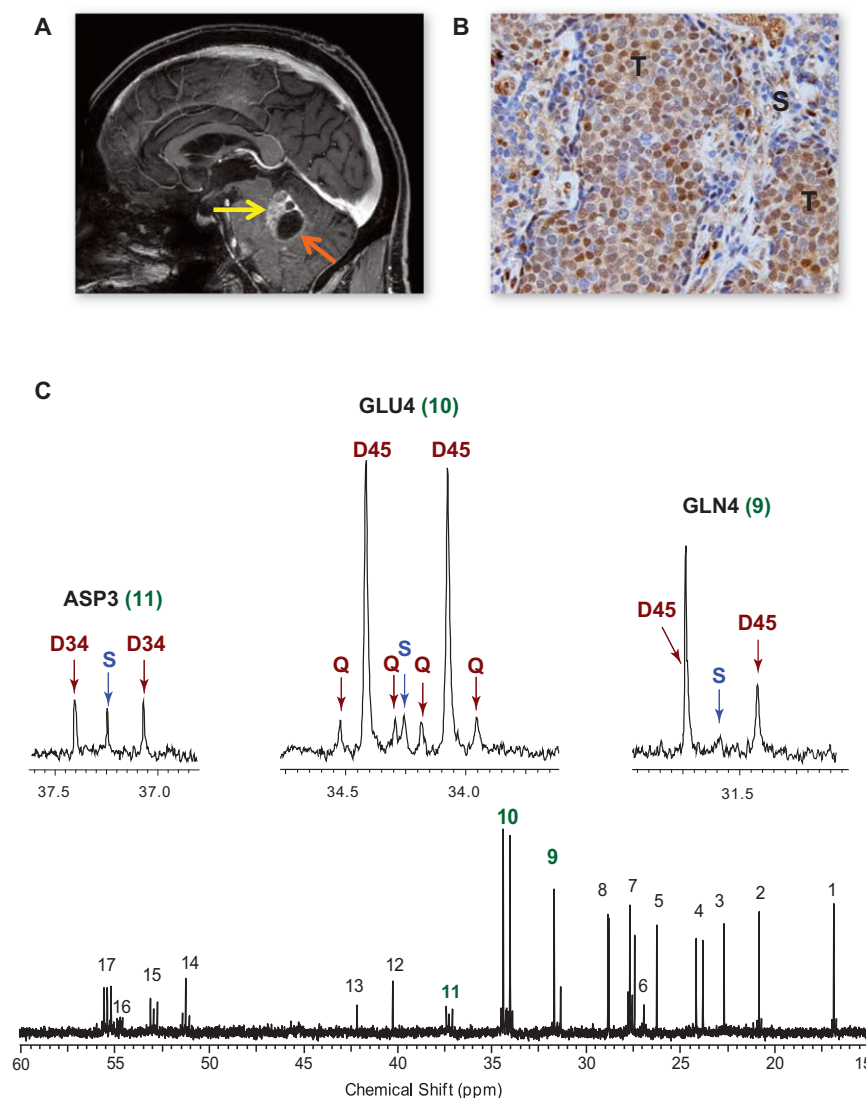


Figure 7. Infusion of [1,2¹³C]Acetate in a Patient with a Non-Small Cell Lung Cancer Brain Metastasis

(A) Preoperative MRI sagittal image shows a gadolinium-enhancing tumor in the left cerebellum (yellow arrow) with a cystic component (orange arrow).

(B) Moderate ACSS2 immunoreactivity in the tumor (T) with a lack of staining in the surrounding stroma (S). Scale bar, 10 μ m.

(C) ¹³C-NMR spectrum with GLU4, GLN4, and ASP3 insets. Abbreviations are the same as in Figure 2. Chemical shift assignments are the same as in Figure 6, with the following additions: 11, aspartate C3; 13, glycine C2; 14, alanine C2; 15, aspartate C2.

See also Figure S4 for ¹³C-spectra from two additional patient tumors.

system than glial cells, the finding that labeled glutamate from ¹³C-acetate increased markedly in a time-dependent manner in ACSS2 null MEFS when the WT ACSS2 was reintroduced into the cells is consistent with the wide range of tumor cell types that are able to oxidize acetate.

Recent evidence suggests that in order for tumor cells to adapt to the brain micro-environment, they must undergo unique genetic adaptations (Valiente et al., 2014). In view of this formidable challenge, identification of a potentially druggable target, such as ACSS2, which appears to have little known function in normal cells, offers unique advantages in therapy development. The current work and that of Comerford et al. (2014) underscores the value of in vivo studies in cancer metabolism. Having orthotopic tumor models

expression of ACSS2 in GBM when compared to the lower grade gliomas supports the assertion that upregulation of this enzyme is linked to an increase in acetate oxidation by the tumor. Higher expression of ACSS2 is associated with shorter survival among the patients with grade II and III gliomas, potentially identifying tumors that are destined to transform to grade IV (GBM) more rapidly.

GBM neurospheres have been shown to be critically dependent on oxidative phosphorylation, but not on inhibition of glycolysis (Janiszewska et al., 2012). This finding is consistent with our observation that inhibition of ACSS2 in the GBM neurospheres results in loss of clonogenicity and cell death and supports the hypothesis that ACSS2 is important for supporting GBM viability. Moreover, the data from the genetically engineered glioma model show that mutations that drive AKT and ERK pathways (loss of *PTEN* and constitutive *BRAF*^{V600E} activation) converge to increase ACSS2 expression (Figure 5). The most direct link between the enzyme and the cell's ability to oxidize acetate in the CAC was shown in ACSS2 null MEFS. Although this is a different

of both GBM and a wide range of brain metastases that have been vigorously cross-validated with human data will enable rapid in vivo workup of novel small-molecule inhibitors that target acetate metabolism, such as ACSS2. In addition, the feasibility of the ¹³C-nutrient methodology in patients enables a host of follow-up questions to be addressed from this work, including whether the ability to oxidize acetate is a property of early-stage tumors or of the acquisition of metastatic potential. Rapid translation is a critical need for patients suffering from GBM and brain metastases, where prognosis is measured in months.

EXPERIMENTAL PROCEDURES

Generation of HOT Mouse Models of GBM and Brain Metastases

Generation of the HOT models of GBM and brain metastases have been previously described (Marin-Valencia et al., 2012a). Briefly, tumor samples were obtained from patients at UTSW Medical Center at the time of craniotomy and tumor resection after written informed consent under an IRB-approved clinical protocol. All brain metabolic labeling experiments were performed in awake, alert mice following placement of an indwelling jugular venous

catheter. For the isolation of tumor and NT brain regions, the whole brains were cut into 1 mm coronal sections. The sections were then microdissected to minimize contamination of tumor and NT regions.

Primary Cell Cultures

Primary astrocyte cultures were cultured as previously described (Bachoo et al., 2002). They were isolated from genetically engineered, multiallele conditional mice with the following mutations as single or combination ($p53^{fl}$, $PTEN^{fl}$, and $LSL-BRAF^{V600E}$). Primary GBM neurospheres were isolated and maintained under standard neurosphere culture conditions.

Infusions of ^{13}C -Labeled Nutrients and Dissection of Brains and Tumors

Mouse Studies

Tracer studies in the HOT mice were performed as previously described (Marin-Valencia et al., 2012c). Mice were infused with [$1,6\text{-}^{13}\text{C}$] glucose ($1\text{-}^{13}\text{C}$, 99% enriched; $6\text{-}^{13}\text{C}$, 99% enriched; Sigma-Aldrich Isotec) and [$1,2\text{-}^{13}\text{C}$] acetate ($1\text{-}^{13}\text{C}$, 99% enriched; $2\text{-}^{13}\text{C}$, 99% enriched; Cambridge Isotope Laboratories) as a bolus of 0.3 mg/g of body weight for each tracer (in 0.3 ml of saline) infused over 1 min, followed by a continuous infusion of 0.0069 mg/g of body weight/min for each tracer (in 0.431 ml of saline) at 150 $\mu\text{l/hr}$ for 150 min. For the experiments using glutamine, mice were infused with [$U\text{-}^{13}\text{C}$] glutamine (95% enriched; Sigma-Aldrich Isotec) as a bolus of 0.28 mg/g of body weight for the tracer (in 0.3 ml of saline) infused over 1 min, followed by a continuous infusion of 0.005 mg/g of body weight/min (in 0.45 ml of saline) at 150 $\mu\text{l/hr}$ for 180 min.

Human Studies

Patients were enrolled in a UTSW IRB-approved protocol to infuse ^{13}C -isotopes (glucose and acetate). For [$1,2\text{-}^{13}\text{C}$]acetate infusion, patients were dosed with 6 mg/kg/min for 5 min, followed by 3 mg/kg/min for 2–3 hr. Patient recruitment and consent, as well as blood and tumor sampling, have been previously described (Maher et al., 2012).

^{13}C -NMR Spectroscopy

Proton-decoupled ^{13}C -spectra of tumor and NT brain extracts were acquired at 150 MHz for ^{13}C on Agilent VNMRs Direct Drive Console using 3 mm broadband NMR probe (Agilent Technologies). Various ^{13}C resonances were assigned based on chemical shift position referenced to the lactate C3 singlet at 20.8 ppm. Relative peak areas of the multiplets were obtained using ACD NMR Processor as previously described (Maher et al., 2012; Malloy et al., 1987, 1990a).

IHC

Paraffin-embedded GBM and brain metastasis specimens were obtained from the Division of Neuropathology at UTSW. Formalin-fixed paraffin-embedded (FFPE) 4 μm sections were used for IHC. The primary antibodies and their dilutions were as follows: rabbit monoclonal ER (Ventana, cat# 790-4325, 1:100 dilution), rabbit monoclonal PR (Ventana, cat# 790-4296, 1:200 dilution), rabbit monoclonal HER-2/neu (Ventana, cat# 790-2991, 1:200 dilution), and rabbit polyclonal ACSS2 (Cell Signaling, cat# 3658S, 1:200 dilution).

RNA Isolation, cDNA Preparation, and Quantitative PCR

RNA was isolated from cells in culture using an RNA Isolation Kit (QIAGEN RNeasy Mini Kit), following the manufacturer's protocol. cDNA was generated using the iScript cDNA Synthesis Kit (BioRad). Quantitative real-time PCR using iTaq SYBR Green Supermix with ROX (Bio-Rad) was performed on a StepOnePlus Real-Time PCR Systems (Applied Biosystems).

Molecular Analysis

DNA was prepared in accordance with standard methods, and quantitative PCR for *EGFR*, *ALK*, *BRAF V600E*, *PTEN*, *p16*, and *p19* was done with standard primer sets for these genes.

Plasma Acetate Measurement

Plasma acetate was measured using standard manufacturer protocols (Biovision Colorimetric Assay Kit, cat# K658). Briefly, 1 μl of the diluted clear filtrate plasma was used for experimental samples. The reaction was per-

formed at room temperature for 40 min, and an optical reading was taken at 450 nm using spectrophotometer. The unknown sample concentration was calculated from the standard curve.

Metabolite Extraction and Liquid Chromatography-Tandem Mass Spectrometry

Cells were seeded on 6 cm plate to 70%–80% confluency. After two washes in PBS, cells were cultured with glutamine starvation medium (Dulbecco's modified Eagle's medium without glucose and glutamine [cat # D5030, Sigma-Aldrich], 3.7 g/l sodium bicarbonate, 5.9 g/l HEPES, 20 mM glucose, 10% FBS), 2 mM sodium [$1,2\text{-}^{13}\text{C}$]acetate (cat# CLM-440-1; Cambridge Isotope Laboratories) was added to the cells, and metabolites were extracted 0, 0.5, 1, 2 hr after acetate addition. After washing with ice-cold PBS, 500 μl of ice-cold 50% high-performance liquid chromatography-grade methanol was added to the cells. The plate was floated on liquid nitrogen immediately until the methanol was frozen. The plate was removed from the liquid nitrogen and scraped until the methanol was thawed. The suspension was then passed through ten freeze/thaw cycles using a bead beater and liquid nitrogen and then centrifuged at 16,000 g at 4°C for 10 min. Supernatants were transferred to a new tube and dried down. Metabolites were resuspended in 140 μl TBA buffer (5 mM tributylammonium acetate [pH 5.0]) and vortexed for 15 min. The solution was placed on ice for 30 min and then centrifuged at max speed for 5 min. The supernatants were transferred to a new tube, placed on ice for 30 min, and then centrifuged at max speed for 5 min. The supernatants were then processed for liquid chromatography-tandem mass spectrometry analysis as described previously using a water/methanol gradient with TBA as the ion pairing agent. A negative mode method targeting labeled glutamate was established based on examining the pattern of incorporation of labeled acetate carbons into glutamate, via entry of cytosolic acetyl-CoA into the TCA cycle. Analysis of the MS/MS fragmentation pattern for glutamate established two MRM pairs (148/103 and 148/130) for detection of [$1,2\text{-}^{13}\text{C}$]acetate conversion to M+2 glutamate (Tu et al., 2007).

Retroviral Studies

ACSS2-shRNA-expressing retroviral plasmid with a puromycin selection gene was used for knockdown studies. We used a scrambled shRNA as control. Retroviral-containing media was obtained from transfected 293T cells and added to GBM cells, followed by puromycin selection.

Statistical Analysis

All pooled results are reported as mean \pm SEM. For the evaluation of the acetate-to-glucose ratio between tumor and NT brain, a paired t test was used, and results are reported as mean \pm SEM.

For ACSS2 IHC evaluation, the immunostained TMA slides were scored manually by assigning to each core a value for ACSS2 staining intensity on a scale of 0–3 and a value representing the proportion of tumor cells staining on a scale of 0%–100%. These two values (intensity and percent positive cells) were then multiplied to obtain a histoscore (range, 0–300), which was used in further analyses. ACSS2 histoscores were dichotomized to high and low at the median. Kaplan-Meier survival curves were generated and compared with the Mantel-Cox (log rank) test using GraphPad Prism (v. 6.01) software (GraphPad).

For further details, please refer to the Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.025>.

AUTHOR CONTRIBUTIONS

The study was conceived by E.A.M. and R.M.B. T.M. managed all the HOT lines, infusion experiments, and tissue samples. V.V. did the mouse cannulations and infusions, as well as performed all the IHC and analysis of TMAs. V.V., S.S., S.N., and D.K.S. did the orthotopic injections. D.K.S. performed

the ACSS2 expression studies. K.P. did the NMR runs and spectral analysis. Z.H. and Z.K. analyzed plasma acetate. K.J.H. and C.F. developed the TMA, and K.J.H. performed the analysis. B.P.T. performed the MEF experiments. S.G.P. generated the GBM cultures and performed the shRNAi experiments. B.E.M. and S.B. did the patient surgical procedures and obtained tumor tissue. R.J.D., K.P., B.P.T., E.A.M., and R.M.B. performed the integrated analysis and wrote the paper.

ACKNOWLEDGMENTS

We thank Craig Malloy and Dean Sherry for analytical assistance with the NMR spectral data, critical reading of the manuscript, and helpful discussions. We thank Sarah McNeil and Christie Sheppard for significant contributions to the patient studies, Jessica Sudderth for analyzing plasma glucose enrichment, and Dinesh Ramesh for assistance with figure preparation. This work was supported by grants from the Cancer Prevention Research Institute of Texas (RP101243 to R.J.D., E.A.M., and R.M.B.), by grants from the NIH (5R01CA154843 to E.A.M.; R01 CA157996 to R.J.D.), by an NIH Resource Grant (EB015908), and by a Simmons Cancer Center NIH support grant (5P30 CA142543-04). This work was also supported by philanthropic funds from The Annette G. Strauss Center for Neuro-Oncology, The Miller Family Fund in Neuro-Oncology, The Gladie Jo Salvino Fund for Glioblastoma Research at UT Southwestern Medical Center, and the Kenny Can Foundation in Dallas, Texas.

Received: March 24, 2014

Revised: August 5, 2014

Accepted: November 17, 2014

Published: December 18, 2014

REFERENCES

- Bachoo, R.M., Maher, E.A., Ligon, K.L., Sharpless, N.E., Chan, S.S., You, M.J., Tang, Y., DeFrances, J., Stover, E., Weissleder, R., et al. (2002). Epidermal growth factor receptor and Ink4a/Arf: convergent mechanisms governing terminal differentiation and transformation along the neural stem cell to astrocyte axis. *Cancer Cell* 1, 269–277.
- Bartnik-Olson, B.L., Oyoyo, U., Hovda, D.A., and Sutton, R.L. (2010). Astrocyte oxidative metabolism and metabolite trafficking after fluid percussion brain injury in adult rats. *J. Neurotrauma* 27, 2191–2202.
- Brennan, C.W., Verhaak, R.G., McKenna, A., Campos, B., Nounshmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al.; TCGA Research Network (2013). The somatic genomic landscape of glioblastoma. *Cell* 155, 462–477.
- Cloughesy, T.F., Cavenee, W.K., and Mischel, P.S. (2014). Glioblastoma: from molecular pathology to targeted treatment. *Annu. Rev. Pathol.* 9, 1–25.
- Comerford, S.A., Huang, Z., Du, X., Wang, Y., Cai, L., Witkiewicz, A.K., Walters, H., Tantawy, M.N., Fu, A., Manning, H.C., et al. (2014). Acetate dependence of tumors. *Cell* 159, this issue, 1591–1602.
- Deelchand, D.K., Nelson, C., Shestov, A.A., Uğurbil, K., and Henry, P.G. (2009). Simultaneous measurement of neuronal and glial metabolism in rat brain in vivo using co-infusion of [1,6-¹³C₂]glucose and [1,2-¹³C₂]acetate. *J. Magn. Reson.* 196, 157–163.
- Dienel, G.A., and Cruz, N.F. (2006). Astrocyte activation in working brain: energy supplied by minor substrates. *Neurochem. Int.* 48, 586–595.
- Ebert, D., Haller, R.G., and Walton, M.E. (2003). Energy contribution of octanoate to intact rat brain metabolism measured by ¹³C nuclear magnetic resonance spectroscopy. *J. Neurosci.* 23, 5928–5935.
- Fine, H.A. (2014). Bevacizumab in glioblastoma—still much to learn. *N. Engl. J. Med.* 370, 764–765.
- Frost, G., Sleeth, M.L., Sahuri-Arisoylu, M., Lizarbe, B., Cerdan, S., Brody, L., Anastasovska, J., Ghourab, S., Hankir, M., Zhang, S., et al. (2014). The short-chain fatty acid acetate reduces appetite via a central homeostatic mechanism. *Nat. Commun.* 5, 3611.
- Ho, C.L., Yu, S.C., and Yeung, D.W. (2003). ¹¹C-acetate PET imaging in hepatocellular carcinoma and other liver masses. *Journal of nuclear medicine* 44, 213–221.
- Janiszewska, M., Suvà, M.L., Riggi, N., Houtkooper, R.H., Auwerx, J., Clément-Schatlo, V., Radovanovic, I., Rheinbay, E., Provero, P., and Stamenkovic, I. (2012). Imp2 controls oxidative phosphorylation and is crucial for preserving glioblastoma cancer stem cells. *Genes Dev.* 26, 1926–1944.
- Jiang, L., Gulanski, B.I., De Feyter, H.M., Weinzimer, S.A., Pittman, B., Guidone, E., Koretski, J., Harman, S., Petrakis, I.L., Krystal, J.H., and Mason, G.F. (2013). Increased brain uptake and oxidation of acetate in heavy drinkers. *J. Clin. Invest.* 123, 1605–1614.
- Lu-Emerson, C., and Eichler, A.F. (2012). Brain metastases. *Continuum (Minneapolis, Minn.)* 18, 295–311.
- Maher, E.A., Marin-Valencia, I., Bachoo, R.M., Mashimo, T., Raisanen, J., Hatanpaa, K.J., Jindal, A., Jeffrey, F.M., Choi, C., Madden, C., et al. (2012). Metabolism of [U-¹³C]glucose in human brain tumors in vivo. *NMR Biomed.* 25, 1234–1244.
- Malloy, C.R., Sherry, A.D., and Jeffrey, F.M. (1987). Carbon flux through citric acid cycle pathways in perfused heart by ¹³C NMR spectroscopy. *FEBS Lett.* 212, 58–62.
- Malloy, C.R., Sherry, A.D., and Jeffrey, F.M. (1988). Evaluation of carbon flux and substrate selection through alternate pathways involving the citric acid cycle of the heart by ¹³C NMR spectroscopy. *J. Biol. Chem.* 263, 6964–6971.
- Malloy, C.R., Sherry, A.D., and Jeffrey, F.M. (1990a). Analysis of tricarboxylic acid cycle of the heart using ¹³C isotope isomers. *Am. J. Physiol.* 259, H987–H995.
- Malloy, C.R., Thompson, J.R., Jeffrey, F.M., and Sherry, A.D. (1990b). Contribution of exogenous substrates to acetyl coenzyme A: measurement by ¹³C NMR under non-steady-state conditions. *Biochemistry* 29, 6756–6761.
- Marin-Valencia, I., Cho, S.K., Rakheja, D., Hatanpaa, K.J., Kapur, P., Mashimo, T., Jindal, A., Vemireddy, V., Good, L.B., Raisanen, J., et al. (2012a). Glucose metabolism via the pentose phosphate pathway, glycolysis and Krebs cycle in an orthotopic mouse model of human brain tumors. *NMR Biomed.* 25, 1177–1186.
- Marin-Valencia, I., Good, L.B., Ma, Q., Malloy, C.R., Patel, M.S., and Pascual, J.M. (2012b). Cortical metabolism in pyruvate dehydrogenase deficiency revealed by ex vivo multiplet (¹³C NMR) of the adult mouse brain. *Neurochem. Int.* 61, 1036–1043.
- Marin-Valencia, I., Yang, C., Mashimo, T., Cho, S., Baek, H., Yang, X.L., Rajagopalan, K.N., Maddie, M., Vemireddy, V., Zhao, Z., et al. (2012c). Analysis of tumor metabolism reveals mitochondrial glucose oxidation in genetically diverse human glioblastomas in the mouse brain in vivo. *Cell Metab.* 15, 827–837.
- Owonikoko, T.K., Arbiser, J., Zelnak, A., Shu, H.K., Shim, H., Robin, A.M., Kalkanis, S.N., Whitsett, T.G., Salhia, B., Tran, N.L., et al. (2014). Current approaches to the treatment of metastatic brain tumours. *Nat Rev Clin Oncol.*
- Oyama, N., Akino, H., Kanamaru, H., Suzuki, Y., Muramoto, S., Yonekura, Y., Sadato, N., Yamamoto, K., and Okada, K. (2002). ¹¹C-acetate PET imaging of prostate cancer. *Journal of nuclear medicine* 43, 181–186.
- Sagiya, K., Mashimo, T., Togao, O., Vemireddy, V., Hatanpaa, K.J., Maher, E.A., Mickey, B.E., Pan, E., Sherry, A.D., Bachoo, R.M., and Takahashi, M. (2014). In vivo chemical exchange saturation transfer imaging allows early detection of a therapeutic response in glioblastoma. *Proc. Natl. Acad. Sci. USA* 111, 4542–4547.
- Schurr, A. (2001). Neuronal energy requirements. In *The Neuronal Environment: Brain Homeostasis in Health and Disease*, W. Walz, ed. (New Jersey: Humana Press), pp. 25–54.
- Sherry, A.D., Malloy, C.R., Zhao, P., and Thompson, J.R. (1992). Alterations in substrate utilization in the reperfused myocardium: a direct analysis by ¹³C NMR. *Biochemistry* 31, 4833–4837.

- Skutches, C.L., Holroyde, C.P., Myers, R.N., Paul, P., and Reichard, G.A. (1979). Plasma acetate turnover and oxidation. *J. Clin. Invest.* *64*, 708–713.
- Taylor, A., McLean, M., Morris, P., and Bachelard, H. (1996). Approaches to studies on neuronal/glial relationships by ^{13}C -MRS analysis. *Dev. Neurosci.* *18*, 434–442.
- Tollinger, C.D., Vreman, H.J., and Weiner, M.W. (1979). Measurement of acetate in human blood by gas chromatography: effects of sample preparation, feeding, and various diseases. *Clin. Chem.* *25*, 1787–1790.
- Tu, B.P., Mohler, R.E., Liu, J.C., Dombek, K.M., Young, E.T., Synovec, R.E., and McKnight, S.L. (2007). Cyclic changes in metabolic state during the life of a yeast cell. *Proc. Natl. Acad. Sci. USA* *104*, 16886–16891.
- Valiente, M., Obenauf, A.C., Jin, X., Chen, Q., Zhang, X.H., Lee, D.J., Chaff, J.E., Kris, M.G., Huse, J.T., Brogi, E., and Massagué, J. (2014). Serpins promote cancer cell survival and vascular co-option in brain metastasis. *Cell* *156*, 1002–1016.
- Yamamoto, Y., Nishiyama, Y., Kimura, N., Kameyama, R., Kawai, N., Hatakeyama, T., Kaji, M., and Ohkawa, M. (2008). ^{11}C -acetate PET in the evaluation of brain glioma: comparison with ^{11}C -methionine and ^{18}F -FDG-PET. *Mol. Imaging Biol.* *10*, 281–287.

Sirtuin 4 Is a Lipoamidase Regulating Pyruvate Dehydrogenase Complex Activity

Rommel A. Mathias,^{1,2} Todd M. Greco,¹ Adam Oberstein,¹ Hanna G. Budayeva,¹ Rumela Chakrabarti,¹ Elizabeth A. Rowland,¹ Yibin Kang,¹ Thomas Shenk,¹ and Ileana M. Cristea^{1,*}

¹Department of Molecular Biology, Princeton University, Princeton, NJ, 08544, USA

²Department of Biochemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, 3086, Australia

*Correspondence: icristea@princeton.edu

<http://dx.doi.org/10.1016/j.cell.2014.11.046>

SUMMARY

Sirtuins (SIRT5) are critical enzymes that govern genome regulation, metabolism, and aging. Despite conserved deacetylase domains, mitochondrial SIRT4 and SIRT5 have little to no deacetylase activity, and a robust catalytic activity for SIRT4 has been elusive. Here, we establish SIRT4 as a cellular lipoamidase that regulates the pyruvate dehydrogenase complex (PDH). Importantly, SIRT4 catalytic efficiency for lipoyl- and biotinyl-lysine modifications is superior to its deacetylation activity. PDH, which converts pyruvate to acetyl-CoA, has been known to be primarily regulated by phosphorylation of its E1 component. We determine that SIRT4 enzymatically hydrolyzes the lipoamide cofactors from the E2 component dihydrolipoyllysine acetyltransferase (DLAT), diminishing PDH activity. We demonstrate SIRT4-mediated regulation of DLAT lipoyl levels and PDH activity in cells and in vivo, in mouse liver. Furthermore, metabolic flux switching via glutamine stimulation induces SIRT4 lipoamidase activity to inhibit PDH, highlighting SIRT4 as a guardian of cellular metabolism.

INTRODUCTION

Sirtuins (SIRT5) are a family of seven mammalian nicotinamide adenine dinucleotide (NAD⁺)-dependent enzymes that regulate diverse biological processes, including genome regulation, stress response, metabolic homeostasis, and aging (Guarente, 2000; Imai et al., 2000). SIRT5 display widespread subcellular distributions, as SIRT1, SIRT6, and SIRT7 are nuclear, SIRT2 is predominantly cytoplasmic, and SIRT3–5 are mitochondrial (Haigis et al., 2006; Michishita et al., 2005). As all SIRT5 have a conserved deacetylase domain, these enzymes are generally known as lysine deacetylases, acting in opposition to acetyltransferases to remove acetyl-modifications from lysine residues (Imai et al., 2000). However, SIRT5 exhibit varying catalytic efficiencies to this modification. SIRT51–3 display robust deacetylase activity, in contrast to SIRT54–5 that show little to no activity (Haigis et al., 2006; Michishita et al., 2005; Schuetz et al., 2007). Emerging evidence has revealed that several SIRT5 can hydro-

lyze alternative lysine modifications more efficiently than acetyl. Specifically, SIRT5 preferentially desuccinylates and demalonylates protein substrates (Du et al., 2011; Peng et al., 2011), while SIRT6 can hydrolyze long-chain fatty acyl lysine modifications (Jiang et al., 2013). These studies have highlighted the functionally dynamic nature of this family of proteins, which are able to perform different enzymatic reactions and regulate a wide range of cellular processes.

Mitochondrial SIRT3–5 regulate ATP production, apoptosis, and cell signaling (Verdin et al., 2010) through distinct enzymatic functions. SIRT3 is considered to be the major deacetylase of the mitochondria, as SIRT3-deficient mice exhibit significant protein hyperacetylation (Lombard et al., 2007). The desuccinylase activity of SIRT5 was shown to target proteins involved in fatty acid β -oxidation and ketone body synthesis pathways, with SIRT5-deficient mice exhibiting an accumulation of acylcarnitines and a decrease in β -hydroxybutyrate production (Rardin et al., 2013). More recently, SIRT5 was reported to regulate lysine glutarylation levels, thereby modulating the activity of carbamoyl phosphate synthase 1, a critical enzyme in the urea cycle (Tan et al., 2014). In contrast to SIRT3 and SIRT5, SIRT4 enzymatic functions have generally remained more elusive (Newman et al., 2012). SIRT4 has been reported to regulate glutamine metabolism (Csibi et al., 2013; Jeong et al., 2013) and fatty acid oxidation via PPAR- α activity (Laurent et al., 2013a). To date, the enzymatic activity of SIRT4 is largely based on its ability to ADP-ribosylate glutamate dehydrogenase (GLUD1), which regulates amino-acid-dependent insulin secretion (Haigis et al., 2006). The deacetylase activities of SIRT4 have remained less well characterized. Initial studies reported limited deacetylation activity (Lin et al., 2012; Michishita et al., 2005), yet SIRT4 was recently reported to control lipid catabolism through deacetylation of malonyl-CoA decarboxylase (MCD) (Laurent et al., 2013b). Additionally, acetylated SIRT4 substrate candidates have been identified in vitro via peptide microarrays (Rauh et al., 2013) and by screening the activity of recombinant SIRT5 against various acyl-histone peptides (Feldman et al., 2013). Unfortunately, these efforts may have been hampered by difficulty in maintaining soluble and active recombinant SIRT4. Therefore, reconciliation of in vitro enzymatic activities with in vivo biological substrates and downstream physiological functions remains a challenge.

Here, we characterized SIRT4 protein interactions within mitochondria, identifying its association with proteins containing

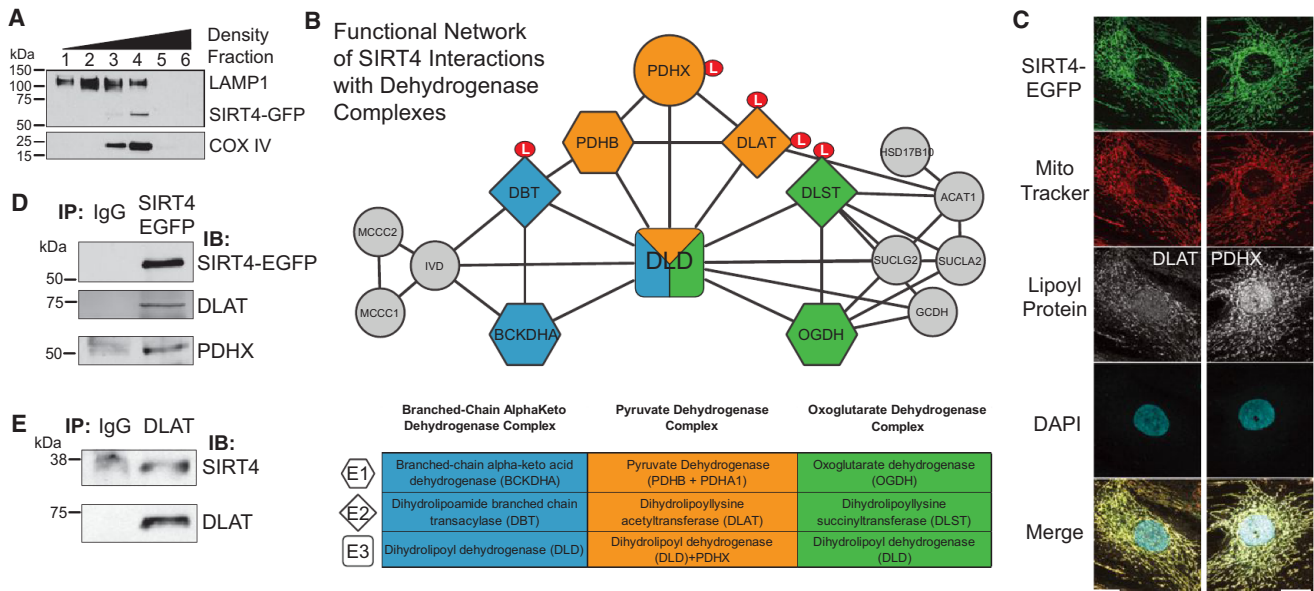


Figure 1. SIRT4 Interacts with the Pyruvate Dehydrogenase Complex

(A) Density gradient-based cellular fractionation of MRC5 cells isolates SIRT4-EGFP with mitochondrial marker COX IV.

(B) Functional network of SIRT4 interactions reveals association with dehydrogenase complexes. The E2 components in each complex (diamonds) contain lipamide modifications (red circle).

(C) SIRT4-EGFP (green) colocalizes with DLAT and PDHX (white) within mitochondria (MitoTracker Red).

(D) Immunoaffinity purification of SIRT4-EGFP coisolates DLAT and PDHX.

(E) Reciprocal immunoaffinity purification of DLAT coisolates endogenous SIRT4 in wild-type fibroblasts. See also Figure S1.

lipoyl and biotinyl modifications. In agreement with this, we demonstrate that SIRT4 removes lipoyl- and biotinyl-lysine modifications more efficiently than acetylations. We discover a physical and functional interaction between SIRT4 and the components of the pyruvate dehydrogenase complex (PDH). PDH is a mitochondrial complex comprised of three catalytic subunits (E1, pyruvate decarboxylase; E2, dihydrolipoilysine acetyltransferase [DLAT]; E3, dihydrolipoil dehydrogenase), a structural subunit (PDH-binding component X [PDHX]) and two regulatory subunits (PDH kinase and PDH phosphatase) (Zhou et al., 2001). The complex catalyzes the decarboxylation of pyruvate to generate acetyl CoA, and links glycolysis to the TCA cycle. Its activity is known to be regulated by phosphorylation of the E1 subunit, phosphorylation that can be also impacted by E1 acetylation (Fan et al., 2014; Jing et al., 2013; Linn et al., 1969; Wieland and Jagow-Westermann, 1969). Here, we show that SIRT4 provides a previously unrecognized, phosphorylation-independent, mechanism of PDH regulation. SIRT4 hydrolyzes lipamide cofactors from the DLAT E2 component of the PDH complex, thereby inhibiting PDH activity. Finally, as glutamine stimulation in rat liver is also known to inhibit the PDH (Häussinger et al., 1982), we investigated whether SIRT4 may play a role in this process. Indeed, we show that glutamine stimulation induces endogenous SIRT4 lipamidase activity, triggering a reduction in both DLAT lipoyl levels and PDH activity. As the PDH controls pyruvate decarboxylation, fueling multiple downstream pathways, our findings highlight SIRT4 as a critical regulator of cellular metabolism.

RESULTS

SIRT4 Interacts with the Three Mitochondrial Dehydrogenase Complexes

To investigate potential cellular substrates of SIRT4, we used proteomics to define its mitochondrial protein interactions. We constructed MRC5 fibroblasts stably expressing SIRT4-EGFP. Using density-based organelle fractionation (coisolation with mitochondrial COX IV, Figure 1A) and direct fluorescence microscopy (colocalization with MitoTracker, Figure 1C and Figure S1A available online), we confirmed its mitochondrial localization. Mitochondria were isolated and the interactions of SIRT4-EGFP were characterized by immunoaffinity purification-mass spectrometry (IP-MS) (Joshi et al., 2013). Interaction specificity was computationally assessed using SAINT (Choi et al., 2011), and 106 significant SIRT4 candidate interactions were identified (Table S1), including the known interactions and substrates, GLUD1, IDE and MLYCD (Ahuja et al., 2007; Haigis et al., 2006; Laurent et al., 2013b). We hypothesized that as yet unrecognized substrates were also identified, and interrogated SIRT4 interactions using bioinformatics to extract enriched metabolic pathways and assemble functional protein networks. Notably, pyruvate metabolism, the TCA cycle, branched-chain amino acid catabolism, and biotin metabolism were significantly enriched pathways (Figure S1). Interaction of SIRT4 with biotin-dependent carboxylases has been reported (Wirth et al., 2013), validating the reliability of our data set. Interestingly, we found that SIRT4 associated with all three of the multimeric mammalian dehydrogenase

Table 1. Determination of In Vitro SIRT4 Kinetics with Acyl-Modified Peptide Substrates

Peptide substrate	Sequence	k_{cat} (s ⁻¹)	K_m (μM)	k_{cat}/K_m (s ⁻¹ M ⁻¹)
H3 K9 Acetyl	KQTARKSTGG <u>W</u> W	ND*	ND* (>2500)	0.083 ± 0.004
H3 K9 Biotinyl		0.0005 ± 0.0001	719 ± 79	0.74 ± 0.05
H3 K9 Lipoyl		0.0019 ± 0.0002	814 ± 163	2.30 ± 0.30
DLAT K259 Acetyl	EIETD <u>K</u> ATIGW	ND*	ND* (>2500)	0.20 ± 0.01
DLAT K259 Lipoyl		0.0018 ± 0.0001	239 ± 51	7.65 ± 1.31
MCD K471 Acetyl	SYLGSK <u>N</u> IKASEW	ND*	ND* (>2500)	0.0064 ± 0.0006

Synthetic peptide sequences are shown containing the modified lysine residues (underlined) as indicated by peptide substrate. When appropriate, k_{cat} , K_m , and k_{cat}/K_m were determined by modeling of kinetic data using the Briggs-Haldane approach (see Figures 2D-E and Materials and Methods). *ND, k_{cat} and K_m could not be determined because v_0 versus [S] was linear. For these cases, k_{cat}/K_m was calculated by linear regression of $v_0/[SIRT4]$ versus [S]. SIRT4 enzyme concentration = 5 μM.

complexes—PDH, oxoglutarate dehydrogenase (OGDH), and branched-chain alpha-keto acid dehydrogenase (BCKDH) (Figure 1B). These complexes occupy discrete positions within the cellular metabolic landscape, regulating TCA cycle activity and amino acid metabolism (Figure S1C). Given its relative prominence within SIRT4 interactions, we focused on PDH. The PDH complex is known to be regulated by reversible phosphorylation of its E1 component (Linn et al., 1969; Wieland and Jagow-Westermann, 1969), with acetylation of E1 also impacting its phosphorylation levels (Fan et al., 2014; Jing et al., 2013). We confirmed that SIRT4-EGFP colocalized (Figure 1C) and immunoprecipitated (Figure 1D) with DLAT and PDH component X (PDHX), the E2 and E3 subunits of PDH, respectively (Figure 1B). Furthermore, in wild-type (WT) human fibroblast cells, we confirmed that DLAT interacts with endogenous SIRT4 by reciprocal IP (Figure 1E).

SIRT4 More Efficiently Catalyzes Removal of Lipoyl- and Biotinyl- Than Acetyl-Lysine Modifications

Given our confirmation of SIRT4 interaction with the PDH components, we pursued their functional relationship. The lipoamide cofactors bound to E2 transferase enzymes (Figure 1B, “L”) are required for PDH activity (Rahmatullah et al., 1990), forming the intermediate S-acetyldihydrolipoyl-lysine in the production of acetyl-CoA. DLAT also has a structural role, constituting the PDH catalytic core. As other mitochondrial SIRT4s can hydrolyze various lysine modifications (Du et al., 2011; Jiang et al., 2013), and as DLAT was prominent in our SIRT4 isolation, we speculated that E2 dehydrogenase components may be biological substrates of SIRT4 and that SIRT4 may directly hydrolyze the lipoamide cofactor. To test this, we screened the in vitro activity of recombinant SIRT4 against differentially-modified synthetic peptides (Table 1). Initially, SIRT4 was incubated with histone H3 Lys9 (H3K9) peptides modified with acetyl-, biotinyl-, or lipoyl-lysine, in the presence or absence of NAD⁺. Following the reaction, the generated unmodified peptides and remaining unreacted substrates were quantified by LC-MS (Figures 2A–2D and Figure S2). SIRT4 only exhibited enzymatic activity in the presence of NAD⁺, shown by the generation of a product peak (P) at ~16.5 min (Figure 2A), corresponding to the unmodified H3K9 (Figure S2A). To compare the relative preference of SIRT4 for the different acyl-lysine peptides, we used extracted ion chro-

matograms to quantify the percentage of unmodified peptide generated for each substrate (Figure 2B). SIRT4 showed the highest potency for removing the lipoyl modification (Figure 2B). The relative amount of unmodified product generated after reaction with SIRT4 was 11% (lipoyl), 3% (biotinyl), and 0.3% (acetyl) (Figure 2B, H3K9 substrates). To show that the enzymatic activity of SIRT4 was required for hydrolysis, we purified a recombinant SIRT4 containing a mutation to the critical residue H161. This histidine, conserved among all SIRT4s, is critical for NAD⁺ and substrate binding (Frye, 1999; Smith and Denu, 2006). In contrast to wild-type SIRT4, deacetylation assays using the catalytically inactive SIRT4 H161Y resulted in no significant activity against any of these acyl-modified substrates (Figure 2C).

SIRT4 Has Superior Lipoamidase Activity for Lipoyl-Modified PDH Peptides

To characterize the putative biological substrates of SIRT4, we tested whether SIRT4 removed lipoamide from DLAT and PDHX peptides (Figures 2B, 2D, S2B, and S2C). SIRT4 showed greater activity toward these substrates than for H3K9, as the proportion of unmodified peptide generated in the presence of NAD⁺ increased to 33% for DLAT and 42% for PDHX (Figure 2B, Lipoyl PDH substrates, and Figure 2D).

Having established SIRT4 enzymatic activity, we next performed steady-state enzyme kinetic assays. This allowed direct comparison of SIRT4's catalytic efficiency for the various acyl-modified peptide substrates (Table 1). Compared to H3K9 acetyl, SIRT4 removed lipoyl and biotinyl H3K9 modifications 28-fold and 9-fold more efficiently, respectively (Figure 3A, Table 1). The DLAT lipoyl peptide displayed a 3.3-fold increase in efficiency compared to H3K9 lipoyl, owing mainly to a decreased K_m (Figures 3A and 3B, Table 1). As a deacetylase, SIRT4 showed slightly greater efficiency toward DLAT acetyl compared to H3K9 acetyl; however, this efficiency was still 38-fold lower than DLAT lipoyl (Figures 3B and S3A). We also compared SIRT4's ability to deacetylate the known biological substrate peptide from MCD (Laurent et al., 2013b) and found that SIRT4 was ~1,270-fold more efficient at hydrolyzing DLAT lipoyl (Table 1, Figure S3A). Altogether, our results demonstrate that SIRT4 has a higher NAD⁺-dependent lipoamidase than deacetylase activity, directly hydrolyzing lipoyl-lysine to generate unmodified lysine (Figure 2E).

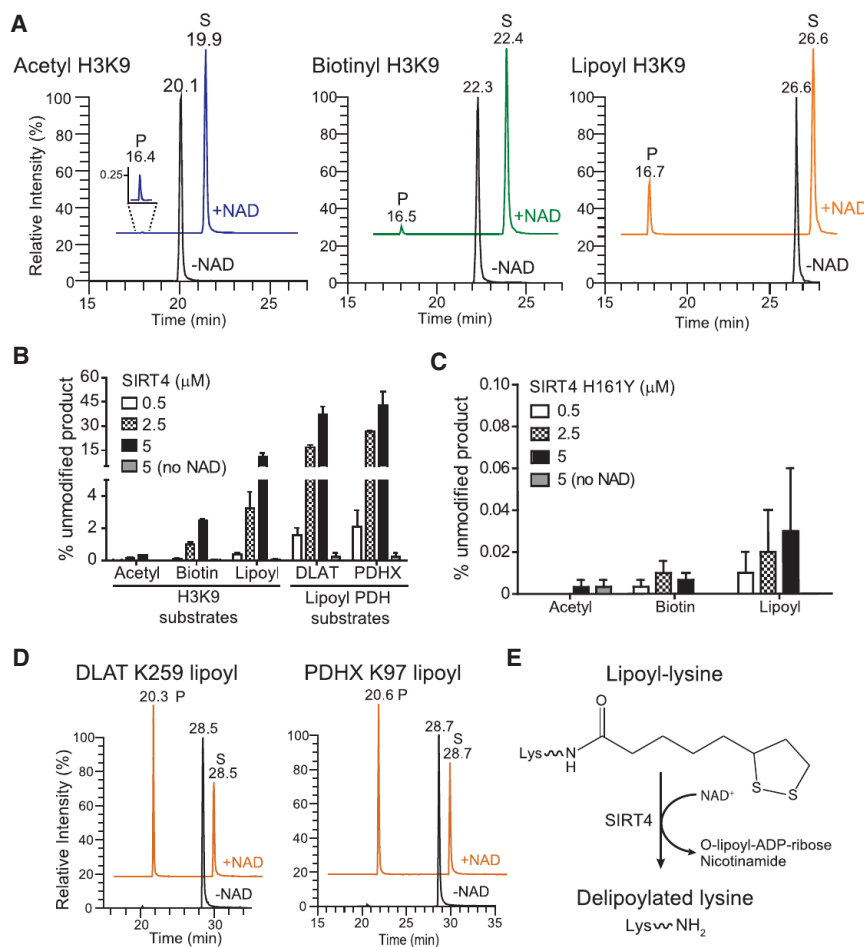


Figure 2. SIRT4 Hydrolyzes Lipoyl-, Biotin-, and Acetyl-Lysine Modifications In Vitro

(A) Recombinant SIRT4 (5 μM) was incubated with various acyl-modified H3K9 peptides (10 μM) with or without NAD (1 mM), and product and residual substrate peptides detected by LC-MS after reaction. Representative extracted ion chromatograms show unreacted acyl-modified H3K9 substrates (S), and unmodified H3K9 products (P, ~16.5 min). +NAD chromatograms are offset for clarity.

(B) The percentage of product (unmodified peptide) formed as a function of increasing concentration of wild-type SIRT4 (±NAD) (mean ± SEM; n = 3).

(C) Same as (B), except product formation from H3K9 substrates after reaction with increasing concentration of catalytically inactive SIRT4 H161Y.

(D) Same as (A), except SIRT4 was incubated with putative biological DLAT and PDHX lipoyl lysine peptides (10 μM). Representative extracted ion chromatograms of unreacted DLAT and PDHX lipoyl peptide substrates (S) and unmodified products (P).

(E) Scheme depicting the NAD⁺-dependent delipoylation of lipoyl-lysine mediated by SIRT4 lipamidase activity. See also Figure S2.

SIRT4 Is the Most Efficient Lipoamidase In Vitro among Mitochondrial Sirtuins

To evaluate the lipoamidase activity for the three known mitochondrial sirtuins (SIRT3–5), we used steady-state enzyme kinetics to compare their ability to hydrolyze lipoyl or acetyl lysine modifications (Figures 3C and 3D). For SIRT5, low but detectable activity was measured for DLAT acetyl (Figure S3B), while no activity was detected for DLAT lipoyl reactions (Figure S3C). As predicted, SIRT3, a robust mitochondrial deacetylase, showed significant enzymatic activity toward DLAT acetyl, while SIRT4 had minimal activity (~800-fold lower) (Figures 3C and S3C). In contrast, although SIRT3 displayed some enzymatic activity toward DLAT lipoyl (Figure 3D), its efficiency was 13-fold lower when compared to DLAT acetyl (Figure 3D versus 3C). Thus, SIRT4 has the highest catalytic efficiency for lipoamide modifications compared to the other mitochondrial SIRTs.

SIRT4 Lipoamidase Activity Diminishes Cellular PDH Lipoamide Levels and Inhibits Its Activity

Given that the lipoamide cofactor is essential for PDH function (Perham, 1991), we examined the impact of elevated SIRT levels on the endogenous cellular activity of PDH by overexpression (OE) of each mitochondrial SIRT in cultured human fibroblasts.

of either SIRT3 or SIRT5 did not alter cellular PDH activity (Figures 4A and S4B), reinforcing the cellular specificity of SIRT4. We further confirmed the direct involvement of SIRT4 activity by showing that PDH activity was not reduced by OE of SIRT4 H161Y (Figures 4A and S4A). Concomitant with the SIRT4-mediated reduction in PDH activity, we observed reduced lipoylation of endogenous DLAT in SIRT4 OE cells, while total DLAT levels remained constant (Figure 4B). DLAT lipoyl levels were not altered in SIRT3 or SIRT5 OE cells. To further characterize the correlation between SIRT4 OE and the decrease in PDH activity and lipoyl levels, we measured phosphorylation of PDH-E1α. Interestingly, we observed reduced phosphorylation at all three sites of E1α in SIRT4 OE cells, while total E1α levels remained constant (Figure 4C). Expression of H161Y SIRT4 did not change phosphorylation levels at any of the E1 sites (Figure 4C), indicating that SIRT4 enzymatic activity is required.

To confirm that the inhibition of PDH activity in cells reflects a direct effect of SIRT4 on the complex, we immunoprecipitated and measured the activity of purified porcine PDH in vitro (Figures 4D and S4C). Purified PDH was treated with pyruvate dehydrogenase phosphatase catalytic subunit 1 (PDP1), which decreased the phosphorylation of all three inhibitory PDH-E1α sites and, as expected, increased PDH activity (Figures 4D

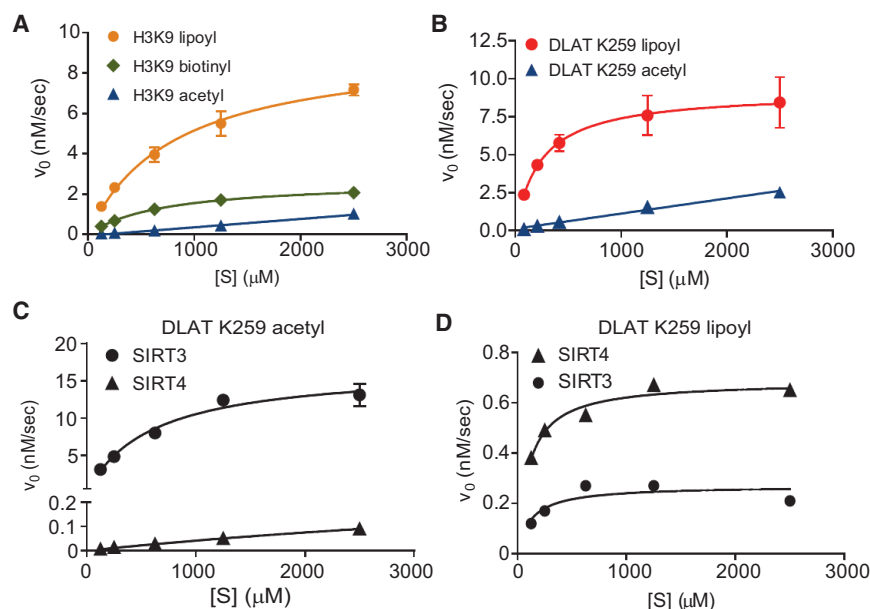


Figure 3. Steady-State Kinetics Reveals SIRT4 Has the Highest Catalytic Efficiency for Lipoyl-Modified Substrates among Mitochondrial SIRTs

(A and B) SIRT4 (5 μ M) initial velocity (v_0) versus substrate concentration [S] for (A) H3K9- and (B) DLAT-acetyl peptides (mean \pm SEM; $n = 3$). v_0 versus [S] were linear for acetyl substrates and were replotted to estimate k_{cat}/K_m (see Figure S3A).

(C and D) Comparison of SIRT3 (0.5 μ M) and SIRT4 (0.5 μ M) initial velocity versus [S] for DLAT K259 (C) acetyl and (D) lipoyl peptide (mean \pm SEM; $n = 3$). SIRT4 v_0 versus [S] was linear for DLAT acetyl and was replotted to estimate k_{cat}/K_m (see Figure S3B). If no error bars are displayed, errors were smaller than the data point size. See also Figure S3.

and S4C). Then, “activated” PDH was treated with either recombinant WT or inactive H161Y SIRT4. Only active SIRT4 was able to attenuate PDH activity, which was not due to increased phosphorylation of E1 (Figure 4D). Overall, these in vitro data with purified PDH support our observations of reduced activity of PDH in SIRT4 OE cells (Figures 4A–4C), and together suggest that reduction in PDH activity occurs in a phosphorylation-independent manner, by SIRT4 directly hydrolyzing lipoylated DLAT.

DLAT contains two lipoyl-lysine residues, K132 and K259. Since western blot analysis only assessed overall lipoyl protein content, we designed an assay using LC-MS/MS selected reaction monitoring (SRM) (Sherrod et al., 2012; Tsai et al., 2012) to measure the effect of SIRT4 on specific lipoylated lysines of endogenous cellular DLAT. Toward this goal, we affinity-purified endogenous DLAT from fibroblast mitochondria and then performed protein digestion using the endoproteinase GluC. Using nontargeted LC-MS/MS the two predicted lipoyl-lysine peptides of endogenous DLAT, containing K132 and K259 residues, were identified (Figures 4E and S4E). We confirmed these results using a synthetic K259 lipoyl peptide, which showed a similar LC retention time and fragmentation pattern as the endogenous K259 peptide (Figure S4F). Additionally, fragmentation of these lipoyl-lysine peptides generated b-ions suitable for relative quantification by targeted MS/MS (Figure 4E, boxes). Using this SRM assay, we measured the effect of SIRT4 OE on the relative levels of endogenous DLAT K132 lipoyl and K259 lipoyl in mitochondria (Figures 4F and S4G). Stable expression of active SIRT4 in fibroblasts reduced levels of DLAT lipoyl at both lysine residues (Figure 4F, left), consistent with our western blotting results (Figure 4B). In contrast, expression of SIRT4 H161Y did not reduce DLAT lipoyl levels. We also analyzed relative lipoyl levels on endogenous DLAT in HEK293 cells transiently transfected with either mCherry (CTL), SIRT4, or SIRT4 H161Y. Only expression of active SIRT4 diminished levels of DLAT lipoyl (Figure 4F,

right), suggesting that SIRT4 reduction of DLAT lipoyl levels were not unique to fibroblasts or an artifact from cell line generation. Altogether, our results demonstrate that the E2 component of the PDH

Glutamine-Stimulation Induces Endogenous SIRT4 Lipamidase Activity and Inhibits PDH Activity

Glutamine stimulation in rat liver is known to cause increased flux through OGDH and decreased flux through PDH, leading to PDH inhibition (Häussinger et al., 1982). Therefore, we investigated whether SIRT4 may play a role in this process. Stimulation of WT fibroblasts with the glutamine supplement glutamax (4 mM) caused a significant time-dependent decrease in PDH activity (Figures 5A, S5A and S5B). Importantly, this reduction in activity was not due to increased levels of inhibitory PDH-E1 phosphorylation relative to unstimulated cells at the same time points (Figure 5B). While steady-state levels of DLAT were unchanged due to glutamax stimulation (Figure 5B), a decrease in DLAT lipoyl levels was observed within 72 hr (Figure 5C). In agreement with these observations, we detected elevated expression of endogenous SIRT4 in cells stimulated with glutamax (Figure 5B). To validate the dependence of PDH inhibition on SIRT4 activity, we measured PDH activity in SIRT4 OE cells stimulated with glutamax. Following 40 hr culture in glutamax, overexpression of active WT SIRT4 triggered pronounced PDH inhibition, in contrast to the H161Y catalytic mutant (Figure S5C). To test the specific involvement of endogenous SIRT4, we generated fibroblasts with knockdown SIRT4 expression using shRNA (Table S2). Effective SIRT4 knockdown was confirmed at the mRNA level (shSIRT4 #1 and #5 achieving > 75% knockdown) (Figure 5D) and at the protein level (Figure 5E). Importantly, SIRT4 knockdown using two different shRNA constructs led to a partial rescue of the glutamax-mediated inhibition of PDH activity (Figures 5F, S5D, and S5F). Finally, to confirm a role for PDH regulation via SIRT4 in vivo, PDH activity was measured in mitochondria purified from the liver of SIRT4 knockout (KO)

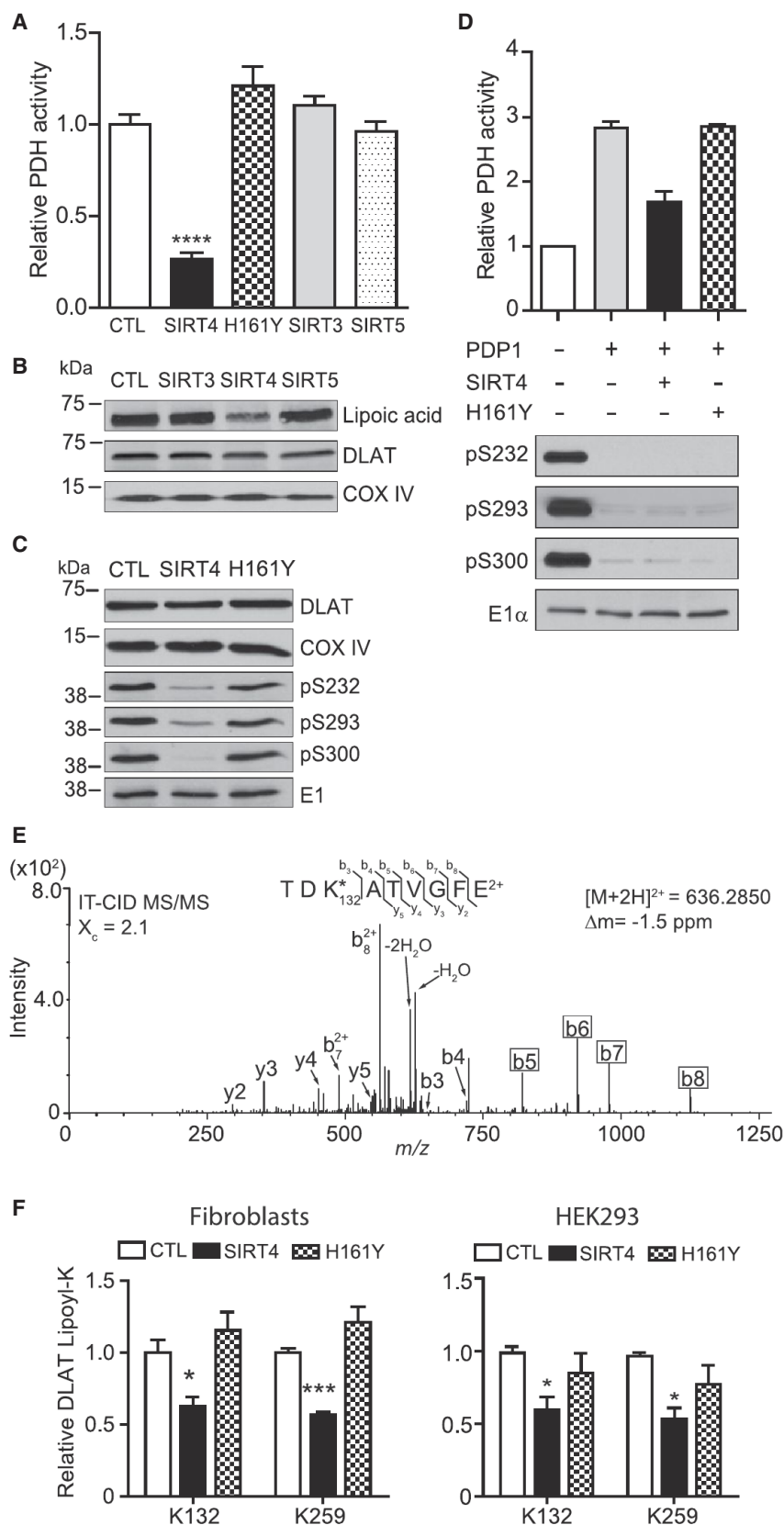


Figure 4. Elevated SIRT4 Expression Decreases the Activity and Lipoylation of the Pyruvate Dehydrogenase Complex in Cultured Cells

(A) PDH activity in fibroblasts expressing mitochondrial SIRT proteins versus GFP-expressing cells (CTL) (mean \pm SEM; $n = 3$ SIRTs 3–5; $n = 5$ GFP; **** $p < 0.0001$) measured by a PDH immunocapture colorimetric assay.

(B) Western blot analysis of endogenous, full-length lipoylated DLAT in cells overexpressing mitochondrial SIRTs. DLAT and COX IV, loading controls.

(C) Western blot analysis of regulatory PDH-E1 α phosphorylation (pS232, pS293, pS300) upon overexpression of SIRT4, catalytically inactive mutant H161Y, or GFP (CTL). E1, loading control.

(D) Relative PDH activity of untreated (control) or “activated” (+pyruvate dehydrogenase phosphatase, PDP1) purified porcine PDH complex incubated with wild-type or H161Y SIRT4. Western blot analysis of PDH-E1 α phosphorylation sites; E1, loading control.

(E) Representative MS/MS spectra of K132 lipoyl peptide detected from endogenous DLAT immunopurified from mitochondria of fibroblasts and digested with endoproteinase GluC. K⁺, reduced and di-carbamidomethylated lipoyl-lysine ($\Delta m = 304$ amu versus unmodified lysine).

(F) SRM quantification of endogenous DLAT lipoyl K132 and K259 in fibroblasts (left) and HEK293 cells (right) (mean \pm S.E.M; $n = 3$, * $p = 0.03$, *** $p = 0.0003$). See also Figure S4.

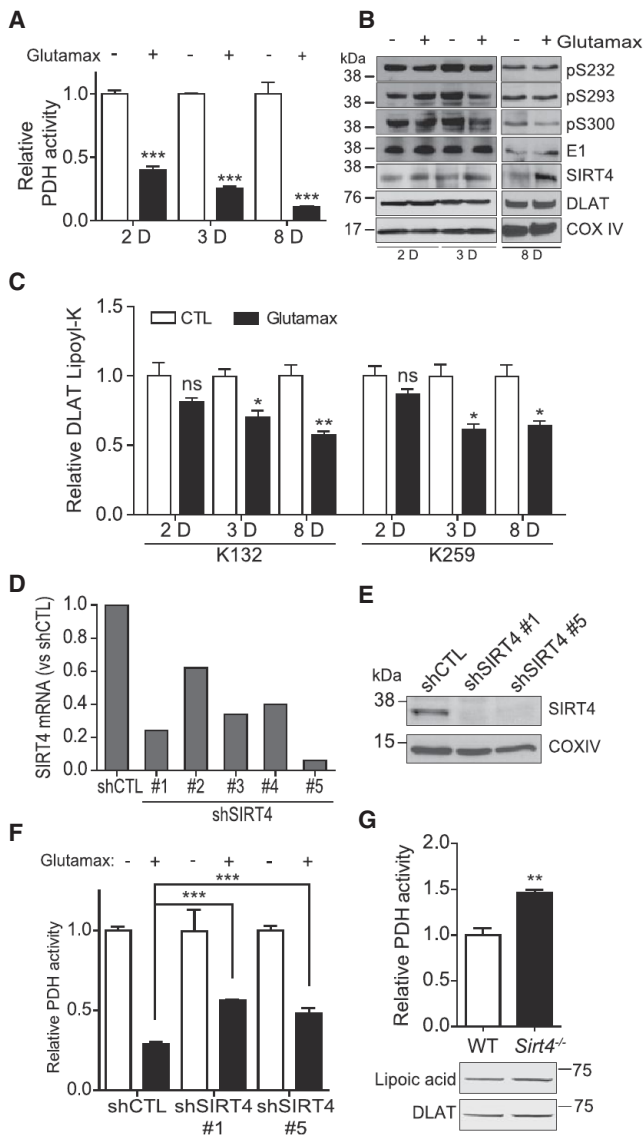


Figure 5. Endogenous SIRT4 Inhibits PDH in Cultured Fibroblasts and In Vivo, in Mouse Liver

(A) PDH activity time course in wild-type fibroblasts stimulated with glutamax (4 mM) for 2, 3, and 8 days, versus unstimulated cells (mean \pm SEM; $n = 4$ 2D and 3D, $p < 0.0001$; $n = 3$ 8D, $p = 0.0007$).

(B) Western blot analysis of regulatory PDH-E1 α phosphorylation sites and total E1 (loading control), and endogenous SIRT4, DLAT, and COX IV (loading control) levels, following glutamax stimulation.

(C) SRM quantification of DLAT lipoyl levels (K132 and K259) in cells stimulated with glutamax versus unstimulated (mean \pm SEM; $n = 3$) for 2D (ns), 3D (* $p = 0.015$), and 8D (** $p = 0.007$, * $p = 0.018$).

(D) Relative SIRT4 mRNA expression measured by qRT-PCR in fibroblast stably expressing either nontargeting control shRNA (shCTL) or one of five different constructs targeting SIRT4 (shSIRT4 #1–5).

(E) Western blot analysis of SIRT4 and COX IV (loading control) from mitochondria purified from fibroblasts expressing shRNA constructs shCTL, shSIRT4 #1, and shSIRT4 #5.

(F) PDH activity in fibroblasts with knockdown levels of endogenous SIRT4 (shSIRT4 #1 or #5, mean \pm S.E.M.; $n = 4$) treated with glutamax (4 mM for 8D), versus control shCTL cells (mean \pm S.E.M.; $n = 7$, *** $p < 0.0001$).

mice. Indeed, we observed elevated PDH activity (Figures 5G and S5E) and DLAT lipoyl levels (Figure 5G) in SIRT4 KO mice relative to control mice. Altogether, these data demonstrate that endogenous SIRT4 is involved in inhibiting PDH activity and DLAT lipoyl levels in the mitochondria of cells and in vivo in mouse liver.

DISCUSSION

Until now, a mammalian cellular lipoamidase has not been characterized. However, our study discovered that SIRT4 can function with this enzymatic capacity in the mitochondria, and that PDH is a biological substrate. We find that, compared to its catalytic efficiency for deacetylation, SIRT4 exhibits far superior enzymatic activity for lipoyl- and biotinyl-lysine modifications. Interestingly, there is precedence for a serum lipoamidase having enzymatic activity for both lipoyl- and biotinyl-lysine modifications (Nilsson and Kågedal, 1993). Importantly, patients with severe serum biotinidase deficiency were observed to exhibit lipoamidase deficiency (Nilsson and Ronge, 1992). Given the serum enzyme was unable to hydrolyze lipoamide from bovine heart PDH (Oizumi and Hayakawa, 1989), it is tempting to speculate that SIRT4 is the mitochondria-specific member of the mammalian class of enzymes that possess both lipoamidase and biotinidase activity and therefore may be unique among these enzymes to liberate lipoate from PDH.

SIRT4 is among the least abundant proteins in the human proteome (PaxDb, www.paxdb.org), and there is only a limited number of human proteins known to be lipoylated. These proteins play critical roles in cellular metabolism and include DLAT, DLST, DBT, PDHX, and GCSH—the first four of which were identified as specific interacting partners of SIRT4 in our IP-MS study. We determined in vitro, in cells, and in a mouse model that SIRT4 regulates overall PDH activity by hydrolyzing the lipoamide cofactors from DLAT. Importantly, we showed that a catalytically active SIRT4 is required for this arm of PDH regulation. Overall, PDH controls pyruvate decarboxylation to generate acetyl-CoA, and DLAT specifically performs the transacetylation reaction, transferring the acetyl-group to Coenzyme A. Our finding that SIRT4 regulates DLAT lipoylation suggests PDH is inhibited via diminished transacetylation. Interestingly, the accessory PDH E3-binding subunit (PDHX) also contains a lipoyl modification. Indeed, we detected lipoylated protein(s) at ~ 50 kDa by western blotting (Figure S4D). This signal may represent individual DLST, DBT, PDHX, or a mixture of these proteins (with similar masses at 49–54 kDa), or other potential SIRT4 substrates. Although this band was reduced in cells upon SIRT4 overexpression (Figure S4D), there was no clear difference in mouse liver from SIRT4 KO when compared to WT mice (Figure S5G). In contrast, DLAT lipoyl levels (~ 70 kDa band) were reduced in cells overexpressing SIRT4 (Figure 4B) and enhanced in SIRT4 KO versus wild-type mice (Figure S5G). This highlights the significance of lipoylated DLAT under SIRT4 null conditions.

(G) PDH activity, lipoyl levels of endogenous DLAT (lipoic acid), and total DLAT levels (DLAT) from mouse liver mitochondria of *Sirt4*^{-/-} mice (mean \pm S.E.M., $n = 3$, ** $p < 0.039$) versus wild-type control ($n = 4$). See also Figure S5.

Additionally, rather than performing a catalytic function, PDHX is known to play a structural role, such as in anchoring the E3 to the E2 (DLAT) subunit (Brautigam et al., 2006; Harris et al., 1997). Thus, should SIRT4, in addition to its impact on DLAT, modulate PDH activity through delipoylation of PDHX, it may involve a structural impairment of PDH. In addition, given our identification of SIRT4 interactions with biotin-dependent decarboxylases and demonstration that SIRT4 also has activity for biotinyl lysine, we predict that SIRT4 regulates decarboxylase activity and potentially other biotin-dependent enzymes and metabolic pathways.

Comparison of steady-state enzyme kinetics is important for determining the catalytic efficiency of an enzyme for particular substrates. However, this has been notoriously difficult for SIRT4 given issues associated with maintaining recombinant protein solubility (Du et al., 2011). Enzyme kinetics have not been reported for SIRT4 substrates prior to this study, and although some *in vitro* activity of SIRT4 toward reduced lipoamide H3K9 peptide has been reported recently (Feldman et al., 2013), the activity was not reproducible, and at very low levels (presumably due to stability issues). We optimized the expression, purification, and storage of SIRT4, which allowed us to perform steady-state kinetics assays and show that SIRT4 has the predominant lipoamidase activity among the mitochondrial sirtuins. Furthermore, SIRT4 catalytic efficiency for lipoylated DLAT is far-superior (1,270-fold) to its previously reported substrate, acetylated MCD, making DLAT the best characterized substrate to date. Importantly, the observed SIRT4 catalytic efficiency and binding constant, K_m , for DLAT-lipoyl is consistent with the cellular lipoyl status of PDH. Specifically, each of the DLAT lipoyl domains is concentrated within PDH at >1 mM (Roche et al., 1993), supporting a cellular role for SIRT4 lipoamidase activity in regulating PDH activity. Indeed, if SIRT4 is actually embedded within the PDH, this finding may help explain the previously reported inability to detect SIRT4 within the mitochondrial matrix by either proteomic profiling (Rhee et al., 2013) or immunofluorescence (www.proteinatlas.org).

Compared to the other mitochondrial SIRTs, our study shows that the cellular lipoamidase activity on DLAT-lipoyl and PDH activity is unique to SIRT4. Elevated expression levels of SIRT3 and SIRT5 in cells did not affect DLAT-lipoyl levels or PDH activity, despite SIRT3 showing some *in vitro* enzymatic activity toward DLAT-lipoyl in kinetic assays. This highlights that the complex architecture and the local cellular environment facilitate protein interactions that may help to define substrate specificity. SIRT4 substrate specificity may also be determined by the size of the active site and amino acids that line the catalytic pocket. For example, structural analysis of the SIRT5 active site defined its preference for lysine substrates bearing negatively charged carboxylates (Du et al., 2011). However, in contrast to all previously characterized SIRT substrates, lipoamide has a sulfur-containing, dithiolane ring. Assuming the SIRT4 active site is near physiological pH, lipoamide will have a neutral charge and may not require extensive charge stabilization. Overall, our results show SIRT4 lipoamidase activity for nonreduced lipoamide *in vitro* (Figure 2); however, the oxidative susceptibility of lipoamide raises the possibility of other *in vivo* substrates. Ultimately, given the current lack of a SIRT4 crystal structure, future studies

will be required to delineate the full range of SIRT4 lipoamide substrate specificity under different cellular states.

Our discovery that SIRT4 inhibits the PDH via direct hydrolysis of the lipoamide cofactor builds on the knowledge accumulated during several decades to provide a new perspective for understanding PDH regulation. PDH activity is understood to be principally inhibited by kinase-dependent phosphorylation of the E1 subunit, whereby phosphorylation is modulated downstream of E1 acetylation (Fan et al., 2014; Jing et al., 2013; Linn et al., 1969; Wieland and Jagow-Westermann, 1969). Our finding that SIRT4 lipoamidase activity can directly impair the function of the complex underscores that PDH regulation is a highly complex process that involves several mechanisms. For instance, we observed that cells with elevated levels of SIRT4 actually have decreased E1 phosphorylation, in conjunction with reduced PDH activity and lipoyl levels. Reduced phosphorylation would normally be expected to activate PDH. Nevertheless, this result is not entirely unexpected and may be partly explained if one considers that PDH kinases are activated via binding to the lipoyl domain (Radke et al., 1993). Therefore, if SIRT4 reduces DLAT lipoamide levels it also triggers an indirect reduction in kinase binding sites, thereby limiting kinase function and causing reduced phosphorylation. Thus, the complex regulation of PDH likely involves the temporal activation of several mechanisms by different stimuli and cellular responses.

Glutamine stimulation in rat liver was shown to inhibit PDH activity by increasing flux through OGDH, accompanied by decreased flux through PDH (Häussinger et al., 1982). Building on this knowledge, we demonstrated that endogenous SIRT4 lipoamidase activity could be induced via glutamine stimulation. Interestingly, after 48 hr of glutamine stimulation the PDH activity decreased by 50%, while the DLAT lipoyl levels decreased to a lesser extent. At subsequent time points, substantial decreases in both PDH activity and DLAT lipoylation were observed. These results suggest that PDH inhibitory mechanisms are temporally regulated. For instance, it is possible that a lipoyl-independent mechanism impacts PDH activity early after glutamine stimulation (up to 48 hr), which is followed by a lipoyl-dependent mechanism that relies on increased SIRT4 levels and lipoamidase activity (72 hr to 8 days). The underlying molecular mechanisms for this temporal regulation are not entirely understood, but may reflect time requirements for increased transcription or translation of SIRT4. Nonetheless, knockdown of SIRT4 in cells led to a partial rescue of the glutamine-mediated inhibition of PDH activity. It remains to be determined whether the lack of a full rescue is due to the function of residual SIRT4 in these knocked-down cells or if another, yet to be identified, PDH inhibition mechanism is simultaneously active. Finally, PDH activity and DLAT lipoyl levels in SIRT4 KO mouse liver were elevated compared to control animals, further supporting the function of SIRT4 in regulating PDH activity *in vivo*.

SIRT4 was initially characterized as an ADP-ribosyltransferase that regulates glutamate dehydrogenase and insulin secretion (Haigis et al., 2006). Studies in insulin producing cells with knockdown SIRT4 expression showed elevated insulin secretion in response to glucose (Ahuja et al., 2007). Our discovery that

SIRT4 inhibits PDH activity has interesting implications for further understanding this phenotype, as insulin secretion is initiated by the rapid utilization of glucose for glycolysis and the consequent entry into the TCA cycle for oxidative phosphorylation (Newgard and McGarry, 1995). Thus, SIRT4 knockdown would help to accelerate this process. SIRT4 has also been reported to regulate fatty acid oxidation via expression of catabolic genes (Laurent et al., 2013a; Nasrin et al., 2010), and by deacetylation of MCD, which inhibits conversion of malonyl-CoA to acetyl-CoA (Laurent et al., 2013b). Our finding that SIRT4 is a lipooamidase is consistent with a function in this metabolic process and highlights PDH as another molecular target that also negatively regulates acetyl-CoA production. SIRT4 has also been reported to be involved in cancer progression. Lung tumors spontaneously develop in SIRT4 KO mice, while expression of SIRT4 represses tumor development in vivo (Csibi et al., 2013; Jeong et al., 2013). Reduced levels of SIRT4 have also been detected in human bladder, breast, colon, gastric, and ovarian carcinoma, relative to normal tissues (Csibi et al., 2013). The aforementioned observations suggest that SIRT4 acts as a tumor suppressor (Zhu et al., 2014), primarily by inhibiting carcinogenesis through repression of glutamine anaplerosis (Jeong et al., 2013). Thus, it is tempting to speculate that this regulation may also involve SIRT4 lipooamidase activity toward OGDH, which contains DLST-lipoyl and feeds 2-oxoglutarate into the TCA cycle.

Taken together, our results identify SIRT4 as a cellular lipooamidase, which regulates the activity of the PDH complex via enzymatic hydrolysis of the lipooamide cofactor. As PDH controls pyruvate decarboxylation, fueling multiple downstream pathways, our findings highlight SIRT4 as a critical regulator of cellular metabolism. We anticipate that these findings will trigger future studies aimed at further characterizing the roles of SIRT4's lipooamidase activity in mitochondrial function in diverse health and disease states.

EXPERIMENTAL PROCEDURES

Cell Culture and Generation of Stable Cell Lines

Human MRC5 fibroblasts and embryonic kidney HEK293 cells were cultured in DMEM containing 10% (v/v) Benchmark fetal bovine serum and 1% penicillin-streptomycin solution. Stable cell lines that express SIRT4-EGFP, SIRT4-EGFP H161Y, or GFP control, as well as stable cell lines with knockdown expression of SIRT4 (using SIRT4-targeting shRNA) or control shRNA were generated as described in Extended Experimental Procedures. SIRT4 expression levels were measured by western blotting, and knockdown efficiency was measured by qRT-PCR and western blotting.

Confocal microscopy

For live microscopy, fibroblasts stably expressing SIRT4-EGFP were imaged on a Leica SP5 confocal microscope using the 63× oil immersion objective. For colocalization studies, SIRT4-EGFP stably expressing fibroblasts were fixed and imaged on a Leica SP5 confocal microscope using the 63× glycerol immersion objective, as described in the Extended Experimental Procedures.

Mitochondrial Isolation and Immunoaffinity Purification of SIRT4-GFP

Purified mitochondria fractions were isolated from fibroblasts (25×10^6) by removing nuclei using centrifugation, and further resolving the resulting crude organelle pellet using a discontinuous OptiPrep gradient, as detailed in Extended Experimental Procedures. SIRT4-EGFP and control EGFP were immunoaffinity purified from mitochondrial fractions using anti-GFP antibodies

and M270 Epoxy Dynabeads, as described in [(Cristea et al., 2005) and in Extended Experimental Procedures.

Proteomic Analysis of SIRT4 and Interacting Protein Partners

SIRT4 immunisolates were analyzed by mass spectrometry-based proteomics using an in-gel digestion approach followed by nanoliquid chromatography (Dionex Ultimate3000) coupled directly to an LTQ-Orbitrap Velos (ThermoFisher Scientific) mass spectrometer operated in data-dependent acquisition mode, as described in Extended Experimental Procedures. The specificity of candidate protein interactions was assessed using the SAINT algorithm, as in (Choi et al., 2011) and in Extended Experimental Procedures. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaino et al., 2014) via the PRIDE partner repository with the data set identifier PXD001447 and DOI 10.6019/PXD001447.

Western Immunoblotting

Proteins were transferred to nitrocellulose membranes, and protein and post-translational modification levels were detected and visualized as described in the Extended Experimental Procedures.

Recombinant Mitochondrial SIRT Proteins

N-terminally truncated human SIRT4 (33-314) was cloned, coexpressed with GroEL and GroES in BL21(DE3) *E. coli*, and purified as described in Extended Experimental Procedures. Recombinant SIRT3 and SIRT5 were purchased from Sigma.

Peptide Synthesis and LC-MS-Based In Vitro Peptide Deacylation Assays

Synthetic peptides were designed (Table 1), synthesized by GenScript, and validated by infusion into an LTQ Orbitrap XL mass spectrometer (ThermoFisher Scientific). The ability of SIRT4 to hydrolyze various acyl-lysine modifications was measured using LC-MS, as described in Extended Experimental Procedures.

HPLC-Based SIRT Kinetic Assays

Kinetic assays for mitochondrial sirtuins and determination of kinetic parameters were performed using HPLC-UV detection as described by others (Du et al., 2011; Jiang et al., 2013) and in the Extended Experimental Procedures. All analyses were performed in a minimum of three biological replicates.

Relative Quantification of Lipoyl-Lysine by Targeted Mass Spectrometry

The relative abundance of lipoyl-lysine-containing DLAT peptides were measured in mitochondria lysates using a selected reaction monitoring (SRM/PRM) full-scan tandem mass spectrometry assay, as described in the Extended Experimental Procedures.

PDH Activity Assay

The activity of the PDH was measured using the Pyruvate Dehydrogenase Enzyme Activity Microplate Assay Kit (Abcam) according to manufacturer's instructions, as detailed in Extended Experimental Procedures. All measurements were performed in at least three biological replicates.

Animal Studies

Experiments in mice were conducted in compliance with Institutional Animal Care and Use Committee (IACUC) of Princeton University. SIRT4 knockout (Jackson Laboratory, Stock number 012756), and control (WT) (Jackson Laboratory, Stock number 002448) adult female mice ($n = 4$) were euthanized, livers excised, and mitochondria isolated as previously described (Rardin et al., 2009) with modifications detailed in Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.046>.

AUTHOR CONTRIBUTIONS

The study was conceived by R.A.M., A.O., T.S., and I.M.C. Experiments performed by R.A.M., T.M.G., H.G.B., and E.A.R. were done in the I.M.C. lab, by A.O. in the T.S. lab, and by R.C. in the Y.K. lab. Studies for identifying SIRT4 substrates were done by R.A.M. Recombinant SIRT4 proteins were purified by A.O. R.A.M. performed MS-based in vitro peptide deacylation assays. R.A.M., A.O., and T.M.G. performed kinetics assays, and R.A.M. and H.G.B. generated the cell lines. H.G.B. performed microscopy and qRT-PCR experiments. T.M.G. performed the SRM-based quantification of lipoylated DLAT, and E.A.R. assessed DLAT lipoyl by western blotting. R.A.M. carried out glutamax experiments and PDH activity assays, and E.A.R. measured SIRT4 mRNA levels. R.C. and Y.K. acquired the mice; R.C. isolated the mice liver and R.A.M., T.M.G. and H.G.B. performed all assays in liver tissue. R.A.M., T.M.G., and I.M.C. prepared the figures and all authors wrote the manuscript.

ACKNOWLEDGMENTS

We thank the PRIDE team for assistance in submitting the mass spectrometry data and interactions to the ProteomeXchange and Intact repositories. The authors are supported by grants from the NIDA (DP1DA026192), NIAID (R21AI102187), and NICHD (R21HD073044) to I.M.C., and NIAID (AI78063) and NCI (CA82396) to T.S., and an ASMS research award to I.M.C. R.A.M. is funded by a NHMRC of Australia Early CJ Martin Fellowship #APP1037043, T.M.G. by an NJCCR postdoctoral fellowship, and H.G.B. and E.A.R. by NSF graduate fellowships.

Received: October 30, 2014

Revised: November 12, 2014

Accepted: November 21, 2014

Published: December 18, 2014

REFERENCES

- Ahuja, N., Schwer, B., Carobbio, S., Waltregny, D., North, B.J., Castronovo, V., Maechler, P., and Verdin, E. (2007). Regulation of insulin secretion by SIRT4, a mitochondrial ADP-ribosyltransferase. *J. Biol. Chem.* 282, 33583–33592.
- Brautigam, C.A., Wynn, R.M., Chuang, J.L., Machius, M., Tomchick, D.R., and Chuang, D.T. (2006). Structural insight into interactions between dihydrolipoamide dehydrogenase (E3) and E3 binding protein of human pyruvate dehydrogenase complex. *Structure* 14, 611–621.
- Choi, H., Larsen, B., Lin, Z.Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.C., and Nesvizhskii, A.I. (2011). SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat. Methods* 8, 70–73.
- Cristea, I.M., Williams, R., Chait, B.T., and Rout, M.P. (2005). Fluorescent proteins as proteomic probes. *Mol. Cell. Proteomics* 4, 1933–1941.
- Csibi, A., Fendt, S.M., Li, C., Poulogiannis, G., Choo, A.Y., Chapski, D.J., Jeong, S.M., Dempsey, J.M., Parkhitko, A., Morrison, T., et al. (2013). The mTORC1 pathway stimulates glutamine metabolism and cell proliferation by repressing SIRT4. *Cell* 153, 840–854.
- Du, J., Zhou, Y., Su, X., Yu, J.J., Khan, S., Jiang, H., Kim, J., Woo, J., Kim, J.H., Choi, B.H., et al. (2011). Sirt5 is a NAD-dependent protein lysine demalonylase and desuccinylase. *Science* 334, 806–809.
- Fan, J., Shan, C., Kang, H.B., Elf, S., Xie, J., Tucker, M., Gu, T.L., Aguiar, M., Lanning, S., Chen, H., et al. (2014). Tyr phosphorylation of PDP1 toggles recruitment between ACAT1 and SIRT3 to regulate the pyruvate dehydrogenase complex. *Mol. Cell* 53, 534–548.
- Feldman, J.L., Baeza, J., and Denu, J.M. (2013). Activation of the protein deacetylase SIRT6 by long-chain fatty acids and widespread deacylation by mammalian sirtuins. *J. Biol. Chem.* 288, 31350–31356.
- Frye, R.A. (1999). Characterization of five human cDNAs with homology to the yeast SIR2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity. *Biochem. Biophys. Res. Commun.* 260, 273–279.
- Guarente, L. (2000). Sir2 links chromatin silencing, metabolism, and aging. *Genes Dev.* 14, 1021–1026.
- Haigis, M.C., Mostoslavsky, R., Haigis, K.M., Fahie, K., Christodoulou, D.C., Murphy, A.J., Valenzuela, D.M., Yancopoulos, G.D., Karow, M., Blander, G., et al. (2006). SIRT4 inhibits glutamate dehydrogenase and opposes the effects of calorie restriction in pancreatic beta cells. *Cell* 126, 941–954.
- Harris, R.A., Bowker-Kinley, M.M., Wu, P., Jeng, J., and Popov, K.M. (1997). Dihydrolipoamide dehydrogenase-binding protein of the human pyruvate dehydrogenase complex. DNA-derived amino acid sequence, expression, and reconstitution of the pyruvate dehydrogenase complex. *J. Biol. Chem.* 272, 19746–19751.
- Häussinger, D., Gerok, W., and Sies, H. (1982). Inhibition of pyruvate dehydrogenase during the metabolism of glutamine and proline in hemoglobin-free perfused rat liver. *Eur. J. Biochem.* 126, 69–76.
- Imai, S., Armstrong, C.M., Kaeberlein, M., and Guarente, L. (2000). Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* 403, 795–800.
- Ishihama, Y., Rappsilber, J., and Mann, M. (2006). Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. *J. Proteome Res.* 5, 988–994.
- Jeong, S.M., Xiao, C., Finley, L.W., Lahusen, T., Souza, A.L., Pierce, K., Li, Y.H., Wang, X., Laurent, G., German, N.J., et al. (2013). SIRT4 has tumor-suppressive activity and regulates the cellular metabolic response to DNA damage by inhibiting mitochondrial glutamine metabolism. *Cancer Cell* 23, 450–463.
- Jiang, H., Khan, S., Wang, Y., Charron, G., He, B., Sebastian, C., Du, J., Kim, R., Ge, E., Mostoslavsky, R., et al. (2013). SIRT6 regulates TNF- α secretion through hydrolysis of long-chain fatty acyl lysine. *Nature* 496, 110–113.
- Jing, E., O'Neill, B.T., Rardin, M.J., Kleinridders, A., Ilkeyeva, O.R., Ussar, S., Bain, J.R., Lee, K.Y., Verdin, E.M., Newgard, C.B., et al. (2013). Sirt3 regulates metabolic flexibility of skeletal muscle through reversible enzymatic deacetylation. *Diabetes* 62, 3404–3417.
- Joshi, P., Greco, T.M., Guise, A.J., Luo, Y., Yu, F., Nesvizhskii, A.I., and Cris- tea, I.M. (2013). The functional interactome landscape of the human histone deacetylase family. *Mol. Syst. Biol.* 9, 672. <http://dx.doi.org/10.1038/msb.2013.1026>.
- Laurent, G., de Boer, V.C., Finley, L.W., Sweeney, M., Lu, H., Schug, T.T., Cen, Y., Jeong, S.M., Li, X., Sauve, A.A., and Haigis, M.C. (2013a). SIRT4 represses peroxisome proliferator-activated receptor α activity to suppress hepatic fat oxidation. *Mol. Cell. Biol.* 33, 4552–4561.
- Laurent, G., German, N.J., Saha, A.K., de Boer, V.C., Davies, M., Koves, T.R., Dephore, N., Fischer, F., Boanca, G., Vaithesvaran, B., et al. (2013b). SIRT4 coordinates the balance between lipid synthesis and catabolism by repressing malonyl CoA decarboxylase. *Mol. Cell* 50, 686–698.
- Lin, H., Su, X., and He, B. (2012). Protein lysine acylation and cysteine succinylation by intermediates of energy metabolism. *ACS Chem. Biol.* 7, 947–960.
- Linn, T.C., Pettit, F.H., and Reed, L.J. (1969). Alpha-keto acid dehydrogenase complexes. X. Regulation of the activity of the pyruvate dehydrogenase complex from beef kidney mitochondria by phosphorylation and dephosphorylation. *Proc. Natl. Acad. Sci. USA* 62, 234–241.
- Lombard, D.B., Alt, F.W., Cheng, H.L., Bunkenborg, J., Streeper, R.S., Mostoslavsky, R., Kim, J., Yancopoulos, G., Valenzuela, D., Murphy, A., et al. (2007). Mammalian Sir2 homolog SIRT3 regulates global mitochondrial lysine acetylation. *Mol. Cell. Biol.* 27, 8807–8814.
- Michishita, E., Park, J.Y., Burneski, J.M., Barrett, J.C., and Horikawa, I. (2005). Evolutionarily conserved and nonconserved cellular localizations and functions of human SIRT proteins. *Mol. Biol. Cell* 16, 4623–4635.
- Nasrin, N., Wu, X., Fortier, E., Feng, Y., Bare', O.C., Chen, S., Ren, X., Wu, Z., Streeper, R.S., and Bordone, L. (2010). SIRT4 regulates fatty acid oxidation and mitochondrial gene expression in liver and muscle cells. *J. Biol. Chem.* 285, 31995–32002.
- Newgard, C.B., and McGarry, J.D. (1995). Metabolic coupling factors in pancreatic beta-cell signal transduction. *Annu. Rev. Biochem.* 64, 689–719.

- Newman, J.C., He, W., and Verdin, E. (2012). Mitochondrial protein acylation and intermediary metabolism: regulation by sirtuins and implications for metabolic disease. *J. Biol. Chem.* **287**, 42436–42443.
- Nilsson, L., and Kågedal, B. (1993). Co-purification of human serum lipoamidase and biotinidase: evidence that the two enzyme activities are due to the same enzyme protein. *Biochem. J.* **291**, 545–551.
- Nilsson, L., and Ronge, E. (1992). Lipoamidase and biotinidase deficiency: evidence that lipoamidase and biotinidase are the same enzyme in human serum. *Eur. J. Clin. Chem. Clin. Biochem.* **30**, 119–126.
- Oizumi, J., and Hayakawa, K. (1989). Liberation of lipoate by human serum lipoamidase from bovine heart pyruvate dehydrogenase. *Biochem. Biophys. Res. Commun.* **162**, 658–663.
- Peng, C., Lu, Z., Xie, Z., Cheng, Z., Chen, Y., Tan, M., Luo, H., Zhang, Y., He, W., Yang, K., et al. (2011). The first identification of lysine malonylation substrates and its regulatory enzyme. *Mol. Cell. Proteomics* **10**, 012658. <http://dx.doi.org/10.1074/mcp.M1111.012658>.
- Perham, R.N. (1991). Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: a paradigm in the design of a multifunctional protein. *Biochemistry* **30**, 8501–8512.
- Radke, G.A., Ono, K., Ravindran, S., and Roche, T.E. (1993). Critical role of a lipoyl cofactor of the dihydrolipoyl acetyltransferase in the binding and enhanced function of the pyruvate dehydrogenase kinase. *Biochem. Biophys. Res. Commun.* **190**, 982–991.
- Rahmatullah, M., Radke, G.A., Andrews, P.C., and Roche, T.E. (1990). Changes in the core of the mammalian-pyruvate dehydrogenase complex upon selective removal of the lipoyl domain from the transacetylase component but not from the protein X component. *J. Biol. Chem.* **265**, 14512–14517.
- Rardin, M.J., Wiley, S.E., Naviaux, R.K., Murphy, A.N., and Dixon, J.E. (2009). Monitoring phosphorylation of the pyruvate dehydrogenase complex. *Anal. Biochem.* **389**, 157–164.
- Rardin, M.J., He, W., Nishida, Y., Newman, J.C., Carrico, C., Danielson, S.R., Guo, A., Gut, P., Sahu, A.K., Li, B., et al. (2013). SIRT5 regulates the mitochondrial lysine succinylome and metabolic networks. *Cell Metab.* **18**, 920–933.
- Rauh, D., Fischer, F., Gertz, M., Lakshminarasimhan, M., Bergbrede, T., Aladini, F., Kambach, C., Becker, C.F., Zerweck, J., Schutkowski, M., and Steegborn, C. (2013). An acetylome peptide microarray reveals specificities and deacetylation substrates for all human sirtuin isoforms. *Nat Commun* **4**, 2327. <http://dx.doi.org/10.1038/ncomms3327>.
- Rhee, H.W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339**, 1328–1331.
- Roche, T.E., Powers-Greenwood, S.L., Shi, W.F., Zhang, W.B., Ren, S.Z., Roche, E.D., Cox, D.J., and Sorensen, C.M. (1993). Sizing of bovine heart and kidney pyruvate dehydrogenase complex and dihydrolipoyl transacetylase core by quasielastic light scattering. *Biochemistry* **32**, 5629–5637.
- Schuetz, A., Min, J.R., Antoshenko, T., Wang, C.L., Allali-Hassani, A., Dong, A.P., Loppnau, P., Vedadi, M., Bochkarev, A., Sternglanz, R., and Plotnikov, A.N. (2007). Structural basis of inhibition of the human NAD⁺-dependent deacetylase SIRT5 by suramin. *Structure* **15**, 377–389.
- Sherrod, S.D., Myers, M.V., Li, M., Myers, J.S., Carpenter, K.L., Maclean, B., Maccoss, M.J., Liebler, D.C., and Ham, A.J. (2012). Label-free quantitation of protein modifications by pseudo selected reaction monitoring with internal reference peptides. *J. Proteome Res.* **11**, 3467–3479.
- Smith, B.C., and Denu, J.M. (2006). Sir2 protein deacetylases: evidence for chemical intermediates and functions of a conserved histidine. *Biochemistry* **45**, 272–282.
- Tan, M., Peng, C., Anderson, K.A., Chhoy, P., Xie, Z., Dai, L., Park, J., Chen, Y., Huang, H., Zhang, Y., et al. (2014). Lysine glutarylation is a protein posttranslational modification regulated by SIRT5. *Cell Metab.* **19**, 605–617.
- Tsai, Y.C., Greco, T.M., Boonmee, A., Miteva, Y., and Cristea, I.M. (2012). Functional proteomics establishes the interaction of SIRT7 with chromatin remodeling complexes and expands its role in regulation of RNA polymerase I transcription. *Mol. Cell. Proteomics* **11**, 60–76.
- Verdin, E., Hirschey, M.D., Finley, L.W., and Haigis, M.C. (2010). Sirtuin regulation of mitochondria: energy production, apoptosis, and signaling. *Trends Biochem. Sci.* **35**, 669–675.
- Vizcaíno, J.A., Deutsch, E.W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., Dienes, J.A., Sun, Z., Farrah, T., Bandeira, N., et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226.
- Wieland, O., and Jagow-Westermann, B. (1969). ATP-dependent inactivation of heart muscle pyruvate dehydrogenase and reactivation by Mg²⁺. *FEBS Lett.* **3**, 271–274.
- Wirth, M., Karaca, S., Wenzel, D., Ho, L., Tishkoff, D., Lombard, D.B., Verdin, E., Urlaub, H., Jedrusik-Bode, M., and Fischle, W. (2013). Mitochondrial SIRT4-type proteins in *Caenorhabditis elegans* and mammals interact with pyruvate carboxylase and other acetylated biotin-dependent carboxylases. *Mitochondrion* **13**, 705–720.
- Zhou, Z.H., McCarthy, D.B., O'Connor, C.M., Reed, L.J., and Stoops, J.K. (2001). The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes. *Proc. Natl. Acad. Sci. USA* **98**, 14802–14807.
- Zhu, Y., Yan, Y., Principe, D.R., Zou, X., Vassilopoulos, A., and Gius, D. (2014). SIRT3 and SIRT4 are mitochondrial tumor suppressor proteins that connect mitochondrial metabolism and carcinogenesis. *Cancer Metab* **2**, 15.

Muscle Spindle Feedback Directs Locomotor Recovery and Circuit Reorganization after Spinal Cord Injury

Aya Takeoka,^{1,2,4} Isabel Vollenweider,^{3,4} Grégoire Courtine,^{3,5} and Silvia Arber^{1,2,5,*}

¹Biozentrum, Department of Cell Biology, University of Basel, 4056 Basel, Switzerland

²Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland

³Brain Mind Institute and Centre for Neuroprosthetics, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

⁴Co-first author

⁵Co-senior author

*Correspondence: silvia.arber@unibas.ch

<http://dx.doi.org/10.1016/j.cell.2014.11.019>

SUMMARY

Spinal cord injuries alter motor function by disconnecting neural circuits above and below the lesion, rendering sensory inputs a primary source of direct external drive to neuronal networks caudal to the injury. Here, we studied mice lacking functional muscle spindle feedback to determine the role of this sensory channel in gait control and locomotor recovery after spinal cord injury. High-resolution kinematic analysis of intact mutant mice revealed proficient execution in basic locomotor tasks but poor performance in a precision task. After injury, wild-type mice spontaneously recovered basic locomotor function, whereas mice with deficient muscle spindle feedback failed to regain control over the hindlimb on the lesioned side. Virus-mediated tracing demonstrated that mutant mice exhibit defective rearrangements of descending circuits projecting to deprived spinal segments during recovery. Our findings reveal an essential role for muscle spindle feedback in directing basic locomotor recovery and facilitating circuit reorganization after spinal cord injury.

INTRODUCTION

Spinal cord injury has an immediate and devastating impact on the control of movement. The origin of motor impairments lies in the physical disconnection of descending pathways from spinal circuits below the lesion, depriving them of synaptic input essential for the generation and regulation of motor output. Despite the failure of severed axons to regenerate at long distance (Ramon y Cajal, 1928; Tello, 1907), partial lesions of the human spinal cord are frequently associated with spontaneous functional improvement (Curt et al., 2008). One of many challenges in restoring motor control after spinal cord injury is to re-establish a sufficient level of task-specific excitability within disconnected local spinal circuits to drive motor neurons caudal to the injury.

Recent studies on incomplete spinal cord injury animal models uncovered some of the mechanisms that may contribute to spontaneous motor recovery (Ballermann and Fouad, 2006; Bareyre et al., 2004; Courtine et al., 2008; Jankowska and Edgley, 2006; Rosenzweig et al., 2010; Zörner et al., 2014). These investigations showed that recovery correlates with the establishment of intraspinal detour circuits. Such alternative pathways through the spared tissue form novel functional bridges across the lesioned spinal segments. At present, circuit-level mechanisms promoting the formation of detour circuits to restore control of movement remain elusive, even though such insight might play a pivotal role in developing interventions that enhance locomotor recovery after spinal cord injury.

Various studies suggest that sensory information plays a critical role in gait control and in locomotor recovery after spinal cord injury (Edgerton et al., 2008; Pearson, 2008; Rossignol et al., 2006; Rossignol and Frigon, 2011; Windhorst, 2007). The most common medical practice used to facilitate motor recovery of paraplegic patients is weight-supported locomotor rehabilitation (Dietz and Fouad, 2014; Knikou and Mummidisetty, 2014; Roy et al., 2012). Repetitive movement during rehabilitative training likely enhances glutamatergic dorsal root ganglia (DRG) sensory feedback, which constitutes the primary extrinsic source of excitation entering the spinal cord below injury to engage local spinal circuits. This interpretation is supported by evidence from animal models in which spinal cord injury coupled to partial or complete elimination of sensory input impairs gait control and locomotor recovery (Bouyer and Rossignol, 2003; Lavrov et al., 2008). However, the DRG neuron subtype promoting locomotor recovery and the mechanisms by which this process takes place are unclear.

Proprioceptive sensory neurons innervate sense organs in the muscle and transmit information about muscle contraction to the spinal cord (Brown, 1981; Rossignol et al., 2006; Windhorst, 2007). Their influence on the activity of central circuits is essential for modulation and adjustment of motor output (Pearson, 2008; Rossignol et al., 2006). Muscle spindle afferents constitute a subset of proprioceptors contacting muscle spindle sense organs. They exhibit the most widespread central projection pattern of all DRG sensory neurons and establish

synaptic contacts with motor neurons and various classes of interneurons implicated in motor control (Brown, 1981; Eccles et al., 1957; Rossignol et al., 2006; Windhorst, 2007). Muscle spindle afferents are thus in a prime position to convey direct excitation to spinal circuits relevant to the regulation of motor behavior, especially under conditions of disconnected descending input.

The zinc-finger transcription factor *Egr3* is expressed selectively by muscle spindle-intrinsic intrafusal muscle fibers, and its mutation results in early postnatal degeneration of muscle spindles (Tourtellotte and Milbrandt, 1998). This defect abolishes normal function of muscle spindle afferents as assessed electrophysiologically (Chen et al., 2002) and leads to gait ataxia (Tourtellotte and Milbrandt, 1998). *Egr3* mutant mice thus represent a genetic model with DRG sensory neuron dysfunction selectively restricted to muscle spindle afferents. They provide an opportunity to investigate how this feedback channel contributes to gait control in intact mice and influences locomotor recovery and circuit reorganization after spinal cord injury.

To address this question, we conducted kinematic analyses in wild-type and *Egr3* mutant mice. Deficiency of muscle spindle feedback did not affect basic motor abilities in intact *Egr3* mutant mice beyond specific gait features. However, lack of muscle spindle feedback severely restricted spontaneous recovery after incomplete spinal cord injury. *Egr3* mutant mice also exhibit a markedly reduced ability to establish descending detour circuits restoring access to spinal circuits below spinal cord injury. We conclude that muscle spindle feedback is a key neuronal substrate to direct circuit rearrangement necessary for locomotor recovery after incomplete spinal cord injury.

RESULTS

Proficient Basic Locomotion in Absence of Muscle Spindle Feedback

We performed high-resolution video recordings to reconstruct hindlimb kinematics in wild-type and *Egr3* mutant mice (Figures 1A and 1B). We focused on task-dependent contributions of muscle spindle input to hindlimb motor control with the aim to establish a baseline to which we could compare the locomotor recovery process after spinal cord injury.

We first assessed hindlimb motor control during basic overground locomotion. Wild-type and *Egr3* mutant mice performed this task with reciprocal activation of flexor and extensor muscles and alternation between left and right hindlimbs (Figure 1B; Movie S1 available online). However, *Egr3* mutant mice exhibited gait ataxia as reported previously (Tourtellotte and Milbrandt, 1998). To characterize gait patterns, we computed >100 kinematic parameters that provide a comprehensive quantification of locomotor features (Figure S1) (Courtine et al., 2008). We subjected all measured parameters to a principal component (PC) analysis (van den Brand et al., 2012) (averaged values of 10–25 step cycles/hindlimb/mouse; $n = 22$ wild-type and $n = 19$ *Egr3* mutants; Figure S2). We then visualized gait patterns in the new space created by PC1–3, where PC1 explained the highest variance (18%) and distinguished the two genotypes (Figure 1C).

The locomotor phenotype observed in *Egr3* mutant mice was limited to distinct gait features represented in PC1 and approximately 65% of all parameters did not correlate with this genotype-specific PC1 (Figure S3A).

To evaluate the ability of *Egr3* mutant mice to adjust gait patterns to changing locomotor velocities, we tested mice during stepping on a treadmill. Both wild-type and *Egr3* mutant mice were capable of stepping across the entire range of tested treadmill speeds (7–23 cm/s; Figures 1D and S3B). PC1 captured adjustment of gait patterns with increasing speed in mice of both genotypes (16% of explained variance; Figures 1D and S3C), whereas PC2 segregated genotypic differences independent of velocity (10% explained variance; Figures 1D and S3C). Electromyogram (EMG) recordings of ankle extensor and flexor muscles revealed that both genotypes showed appropriate speed-dependent adjustments in burst duration (Figure 1D). These findings resonate with work demonstrating that the flexion phase of the step cycle remains constant, whereas the extension phase progressively shortens with increased locomotor speed (Arshavskii et al., 1965; Halbertsma, 1983), a property we now demonstrate to be independent of muscle spindle sensory feedback.

In summary, both wild-type and *Egr3* mutant mice are able to perform basic locomotor tasks proficiently, but mutant mice display specific gait alterations concordant with the previously proposed role of muscle spindle feedback in control and adjustment of locomotion (Pearson, 2008; Rossignol et al., 2006; Windhorst, 2007).

Muscle Spindle Feedback Is Essential for Locomotor Precision Task and Swimming

Next, we tested mice of both genotypes during walking on a horizontal ladder, requiring precision in foot placement. Whereas wild-type mice progressed across the ladder with ease, *Egr3* mutants frequently slipped off or missed rungs, which was reflected in aberrant bouts of EMG activity (Figure 2A; Movie S2). Quantification of foot positioning relative to successive rungs revealed that wild-type mice targeted rungs precisely, whereas *Egr3* mutant mice displayed near-random foot placement (Figure 2B). These findings demonstrate an essential role for muscle spindle feedback circuits in the regulation of accurate foot placement in a locomotor precision task.

Egr3 mutant mice exhibit selective defects of muscle spindle feedback, but other sensory feedback is preserved (Tourtellotte and Milbrandt, 1998). During swimming, afferents from Golgi tendon organs are attenuated due to reduced weight load (Gruner and Altman, 1980). Proprioceptive signaling therefore relies almost exclusively on muscle spindle feedback. We found that during swimming, wild-type mice displayed well-coordinated alternation of left and right hindlimbs with reciprocal activity of ankle flexor and extensor muscles (Figures 2C and 2D). In contrast, *Egr3* mutant mice were unable to keep afloat and showed uncoordinated hindlimb movements with extensive coactivation of antagonistic muscles (Figures 2C and 2D). Together, these findings stress the pivotal function of muscle spindle feedback in the control of swimming, a condition when Golgi tendon organ and cutaneous feedback circuits only play a limited task-related function.

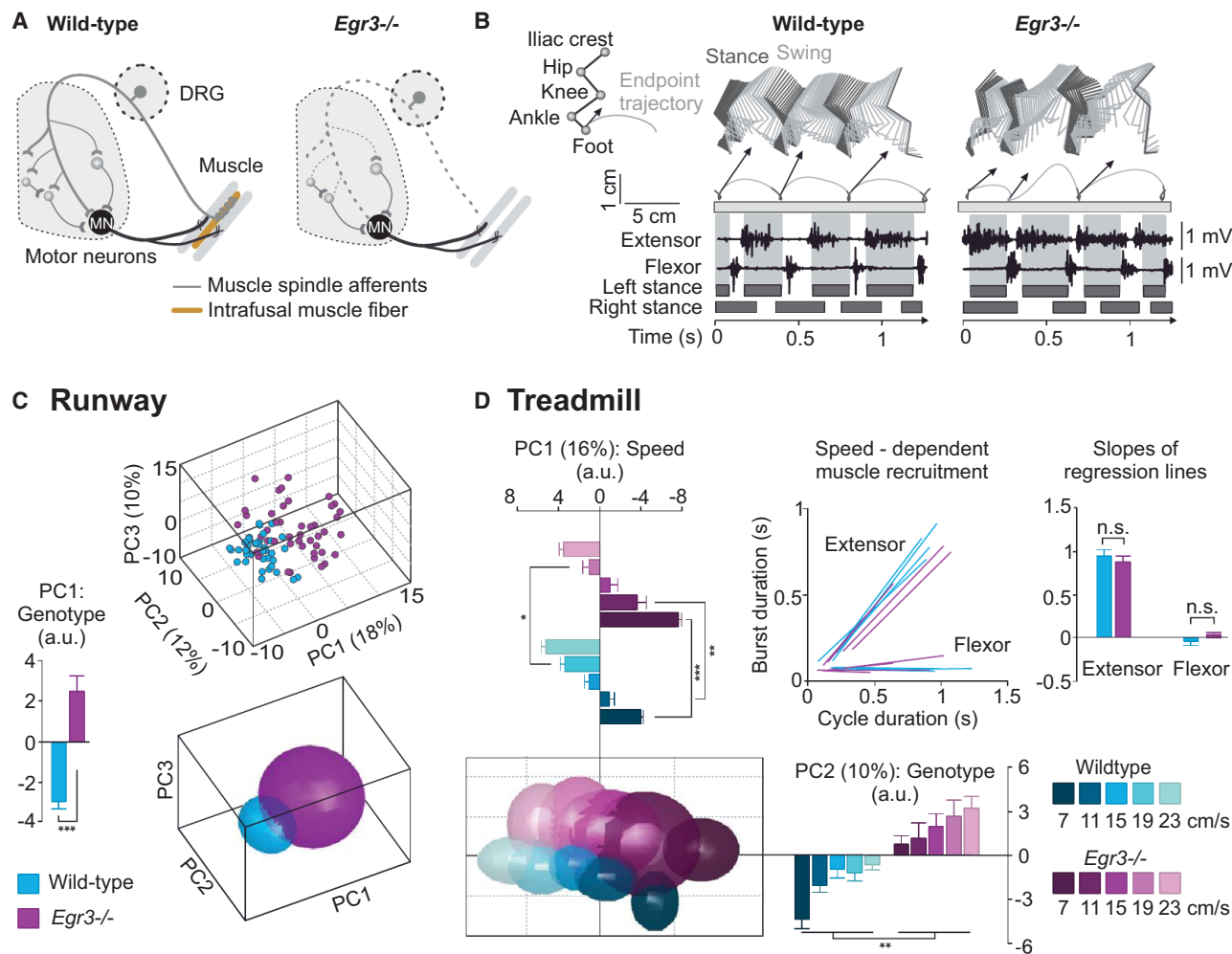


Figure 1. Proficient Basic Locomotion in Absence of Muscle Spindle Feedback

(A) *Egr3* mutation results in selective degeneration of muscle spindles and nonfunctional muscle spindle feedback circuits.

(B) Color-coded stick decomposition of hindlimb movement during three consecutive steps with limb endpoint trajectories and velocity vector at swing onset during basic overground locomotion in both genotypes (EMG activity of an extensor and a flexor muscle displayed below; dark gray bars, stance; empty spaces, swing).

(C) PC analysis was applied on 103 gait parameters measured during overground locomotion (10–25 gait cycles/hindlimb/mouse, $n = 22$ wild-type and $n = 19$ *Egr3* mutants). Gait cycles are represented for each animal and hindlimb (individual dots) in the new space created by PC1–3. Least-squares elliptical fitting (95% confidence) was computed to emphasize differences between genotypes. Histogram plot, mean values of PC1 scores for each genotype.

(D) PC analysis applied on averaged values of 108 gait parameters (15–30 gait cycles/mouse/speed, $n = 10$ for each genotype) measured during stepping on a treadmill at five different speeds (7–23 cm/s). Histogram plots, mean values of PC1 and PC2 scores. Correlation between step-cycle duration and extensor or flexor burst duration. Each regression line was computed separately for a given animal ($n = 4$ for each genotype; 25–30 step cycles/mouse). Histogram plots, slopes of regression lines for extensor and flexor muscles.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; ns, not significant; error bars, SEM; extensor, gastrocnemius medialis; flexor, tibialis anterior; a.u., arbitrary unit. See also Figures S1, S2, and S3 and Movie S1.

Muscle Spindle Feedback Circuits Are Essential for Locomotor Recovery after Injury

The core ability to perform basic locomotion is not disturbed in *Egr3* mutant mice, providing an opportunity to assess the role of muscle spindle feedback circuits in gait control and spontaneous recovery after spinal cord injury. We placed a lateral hemisection injury at the thoracic level (T10) and confirmed lesion completeness upon termination of experiments (Figure 3A). This lesion interrupts descending tracts projecting ipsilaterally to

lesion (ipsilesional hereafter), which normally innervate lumbar segments containing circuits essential for the control of ipsilesional hindlimb muscles (Figure 3A). We performed kinematic analysis at regular intervals after injury to follow locomotor recovery (Figures 3A and 3B).

Three days after injury (acute), both wild-type and *Egr3* mutant mice dragged the ipsilesional hindlimb along the runway as they moved forward (Figure 3B; Movies S3 and S4). Wild-type mice gradually regained locomotor proficiency over the time course

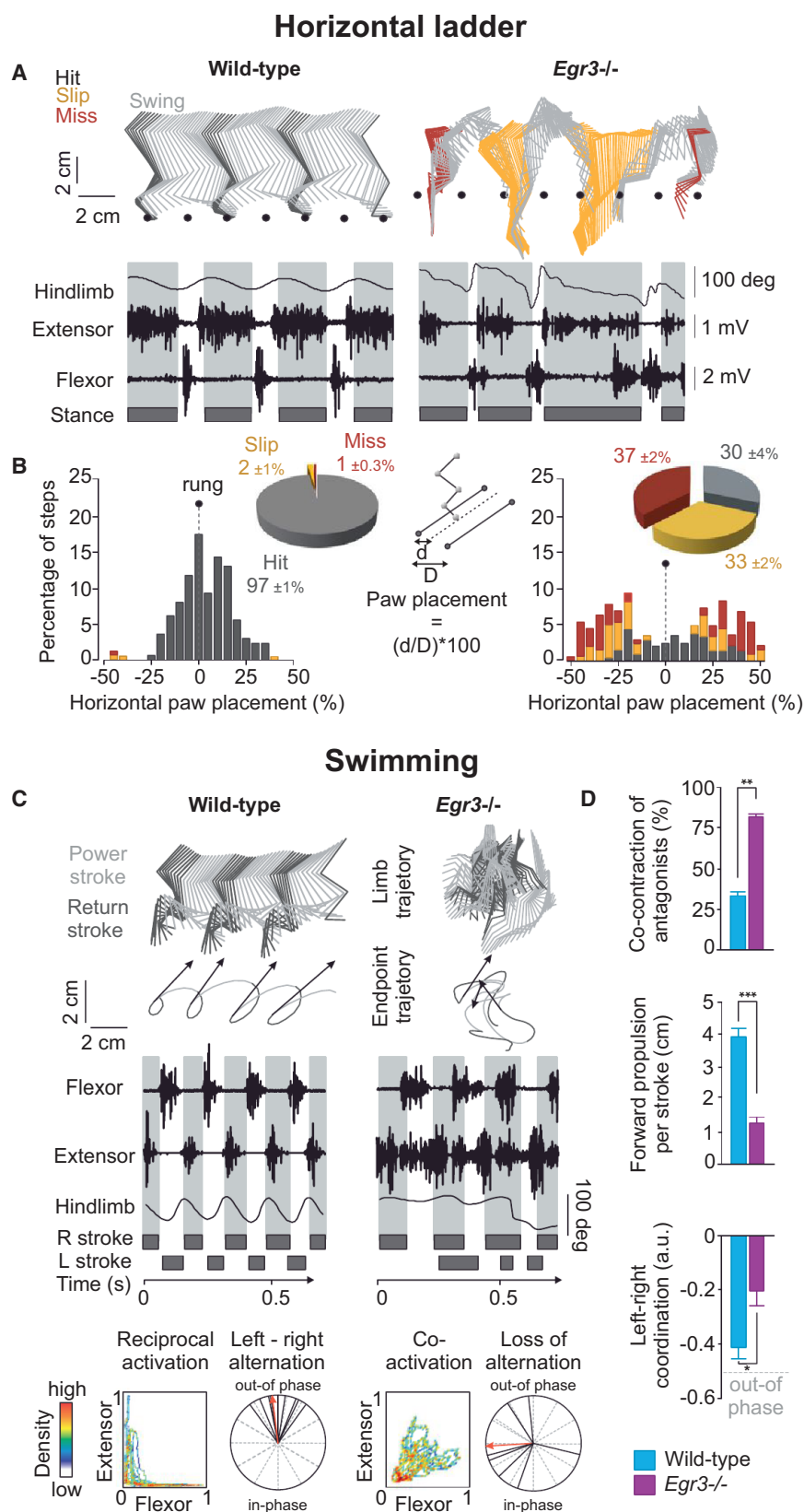


Figure 2. Muscle Spindle Feedback Is Essential for Precision Tasks and Swimming

(A) Stick diagram decomposition of hindlimb movement for a representative wild-type and *Egr3* mutant mouse during crossing of an elevated horizontal ladder with rungs (spacing 2 cm; below: hindlimb oscillation and traces of ankle extensor and flexor muscles for same mouse; dark gray bars, stance).

(B) Bar graph quantifying relative positioning of hindpaws with respect to rung positions. Pie charts summarize total percentage of hits, slips, and misses ($n = 9$ mice per genotype; 259 steps for wild-type and 323 steps for *Egr3* mutant mice).

(C) Stick decomposition of hindlimb movement for a wild-type and *Egr3* mutant mouse during swimming (below: limb endpoint trajectories, limb endpoint velocity vectors at power stroke onset, and raw traces of muscle activity for an extensor and flexor muscle together with hindlimb oscillations; dark gray bars, return stroke). Density plot displays coordination between antagonistic muscles during the represented trial (L-shaped patterns, reciprocal muscle activation; diagonal: continuous coactivation). Polar plot, coordination between left and right hindlimb oscillations (black lines, single gait cycle; red arrow, average of all gait cycles).

(D) Histogram plots report mean values for representative kinematic and muscle activity-related variables extracted from PC analysis ($n = 181$ swim strokes, 10–15 strokes/mouse, $n = 8$ wild-type and $n = 7$ *Egr3* mutant mice) during swimming task.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; error bars, SEM; extensor, vastus lateralis; flexor, tibialis anterior; a.u., arbitrary unit. See also Figures S1 and S2 and Movie S2.

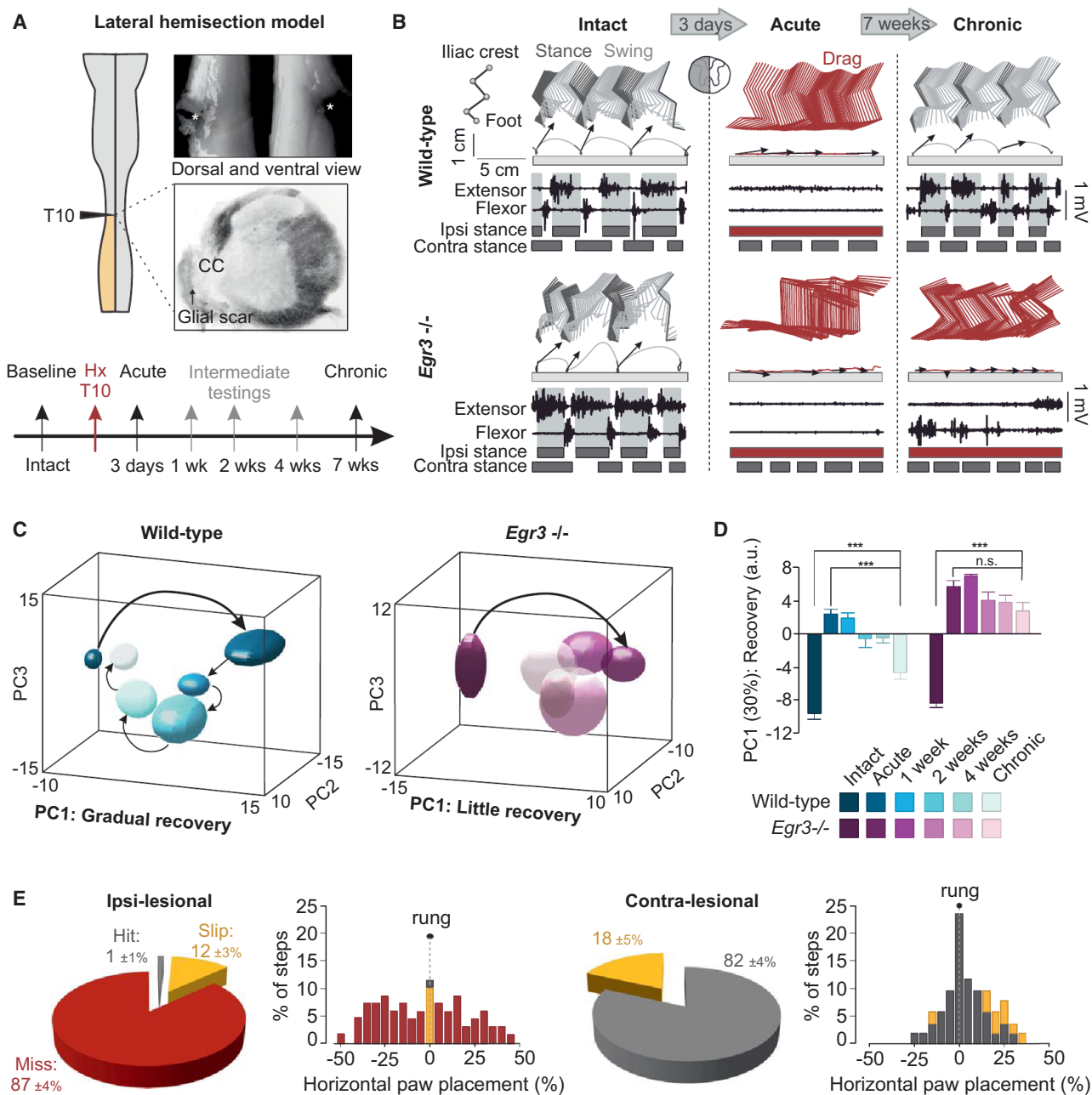


Figure 3. Muscle Spindle Feedback Is Essential to Direct Spontaneous Locomotor Recovery after Lateral Hemisection

(A) Illustration of thoracic lateral hemisection model, including a representative lesion from a dorsal, ventral, and coronal view, and time line of experiment procedures.

(B) For each genotype, a representative stick decomposition of hindlimb movement during basic overground locomotion is shown for intact, acute, and chronic time points for the same mouse (below: concurrent limb endpoint trajectory and velocity vector at swing onset, activity of an extensor muscle, and activity of a flexor muscle; dark gray bars, stance; red bars, dragging).

(C) Representation of gait clusters in PC space for one mouse per genotype during intact stepping and at five different time points postinjury (10–15 steps per time point; 103 parameters per gait cycle).

(D) Histogram plots reporting mean values of PC1 scores measured on all data combined (average of 10–25 steps per time point, 103 parameters per gait cycle, $n = 9$ wild-type mice, $n = 7$ *Egr3* mutant mice).

(E) Bar graph of relative positioning of hindpaws with respect to rung positions for chronically injured wild-type mice ($n = 10$). Pie charts summarize total percentage of hits, slips, and misses ($n = 259$ steps contralesional hindlimb; $n = 147$ steps ipsilesional hindlimb).

* $p < 0.05$; *** $p < 0.001$; ns, not significant; error bars, SEM; extensor, medial gastrocnemius; flexor, tibialis anterior; Hx, hemisection; a.u., arbitrary unit; acute, 3 days postinjury; chronic, 7 weeks postinjury. See also Figures S1, S2, and S4 and Movies S3, S4, and S5.

analyzed. By 7 weeks postinjury (chronic), they regained weight-bearing plantar steps with regular alternation of stance and swing phases of the ipsilesional hindlimb (Figures 3B and S4A; Movies S3 and S4). In contrast, *Egr3* mutant mice still exhibited severe locomotor deficits at the chronic stage (Figures 3B and S4A; Movies S3 and S4).

To quantitatively assess the recovery of ipsilesional hindlimb function, we conducted a PC analysis comparing intact condition to each time point evaluated. We found that PC1 characterized the recovery process (30% explained variance; Figures 3C and 3D). In wild-type mice, time-dependent gait clusters gradually moved toward intact conditions, reflecting the progressive recovery of locomotor function (Figures 3C and 3D). In contrast, gait clusters of *Egr3* mutant mice remained confined in the same PC space through the entire time course evaluated (Figures 3C and 3D). Detailed kinematic analysis revealed that in chronic wild-type mice, 73% of all parameters affected at acute stages improved significantly ($p < 0.05$) and 17% even recovered to levels measured before lesion (Figure S4A). Lack of locomotor recovery in *Egr3* mutant mice was associated with persistent dragging of the ipsilesional hindlimb at all time points (Figures 3B and S4A). In addition, analysis of contralesional hindlimb gait parameters revealed that both wild-type and *Egr3* mutant mice adjust their gait to ipsilesional hindlimb deficiencies similarly and immediately after lesion (Figure S4B). Together, our results demonstrate that defective muscle spindle feedback circuitry severely limits spontaneous locomotor recovery after incomplete spinal cord injury.

Speed Adjustment, but Not Precision Control, Improves in Wild-Type Mice after Injury

Next, we determined the extent to which hemisected wild-type mice regain the capacity to accommodate hindlimb movement to increasing walking speeds and to perform muscle spindle feedback-dependent swimming and ladder precision tasks.

Wild-type mice at chronic stages recovered the ability to walk at the highest speed tested (23 cm/s). After a lack of ipsilesional muscle recruitment at acute stages, the modulation of ankle extensor muscle activity gradually recovered toward intact levels (Figure S4C). In contrast, prolonged paw dragging (Figure S4A) led to increased EMG bursts in ankle flexor muscles after lesion, a feature that only partially recovered at chronic stages (Figure S4C). Wild-type mice regained well-coordinated limb alternation during swimming (Figure S4D) (Zörner et al., 2010), providing further evidence for recovery of basic locomotor features. During precision walking on the horizontal ladder at chronic stages, 87% of ipsilesional hindlimb steps resulted in a complete miss of the targeted rung, and 12% slipped off the rung. In contrast, most steps of the contralesional hindlimb were placed correctly on the rungs (82%) (Figure 3E; Movies S5).

Taken together, these results indicate that after lateral hemisection injury, wild-type mice regain basic locomotor function but only partially recover speed-dependent adaptation and completely fail to recover precise paw placement required for ladder locomotion.

Muscle Spindle-Specific Feedback Needed for Functional Recovery

Contrary to wild-type mice that regained the ability to move their ipsilesional hindlimb after injury, *Egr3* mutants exhibited persistent lack of locomotor control. Because activity-dependent mechanisms contribute to recovery of locomotor function after spinal cord injury (Dietz and Fouad, 2014; Edgerton et al., 2008; Maier and Schwab, 2006), we next measured the degree of spontaneous motility in *Egr3* mutant mice. We monitored home cage activity before injury and at regular intervals after lesion (Figure 4A). Both groups displayed decreased locomotor activity immediately after injury, but there were no significant genotype-related differences in distance covered throughout the recovery process (Figure 4A).

We then asked whether daily administration of monoaminergic receptor agonists known to acutely enhance locomotor output in rodents with severe spinal cord injury (van den Brand et al., 2012) influence the recovery process in *Egr3* mutant mice. We reasoned that despite indistinguishable motility between the two groups after lesion, spinal circuits in *Egr3* mutants may be recruited less efficiently in the absence of functional muscle spindle feedback than in wild-type mice. We found that upon daily agonist administration, *Egr3* mutants still exhibited an overall impediment in locomotor recovery (Figures 4B and 4C). The contribution of individual parameters to the recovery-associated PC1 showed a high correlation between spontaneous and daily drug administered groups for both genotypes (Figure 4D). These results demonstrate that muscle spindle sensory feedback is absolutely essential for directing the process of locomotor recovery after spinal cord injury and cannot be substituted for by daily activation of spinal circuits through pharmacological means.

Muscle Spindle Feedback Promotes Efficient Detour Circuit Establishment around Lesion

Because reorganization of supraspinal and intraspinal descending circuits parallels spontaneous recovery after incomplete spinal cord injury (Bareyre et al., 2004; Courtine et al., 2008; Rosenzweig et al., 2010; Zörner et al., 2014), we asked whether presence of muscle spindle feedback influences these injury-induced circuit reorganization responses. The formation of functional detour circuits relies on the ability of neuronal subpopulations to establish new connections to ipsilesional spinal circuits below the lesion. A predicted hallmark of such neurons is that they must have projections to segments below injury prior to lesion and establish novel synaptic connections after injury.

To identify sources of such neurons, we performed a mapping approach to label neurons with projections to the ipsilesional lumbar spinal cord, by injection of G protein-deficient rabies viruses encoding fluorescent marker proteins (FP) (Rab-FP; Figure 5A) (Wickersham et al., 2007). We analyzed the relative abundance and pattern of marked neurons above lesion in intact, acute, and chronic mice.

We first visualized descending supraspinal projection neurons in intact wild-type and *Egr3* mutant mice. Retrogradely labeled neuron distribution was reminiscent of patterns of mapped premotor brainstem nuclei (Esposito et al., 2014), with most Rab-FP^{ON} neurons located in the magnocellular, followed by pontine,

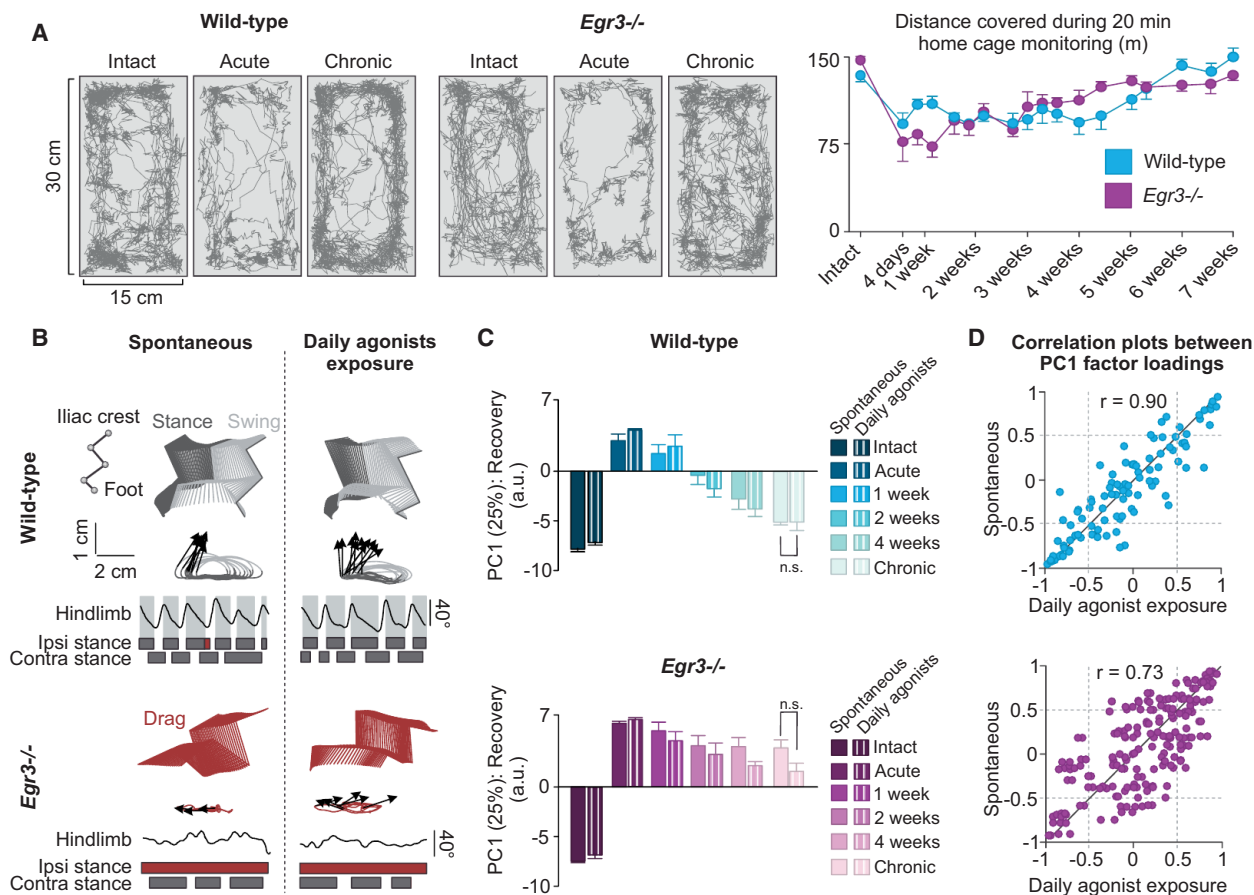


Figure 4. Impact of Activity on Functional Recovery after Spinal Cord Injury

(A) Body trajectories measured during the first 5 min of home cage monitoring for wild-type ($n = 4$) and *Egr3* mutant ($n = 3$) mice. Quantification of distance covered during 20 min period at intact condition and throughout recovery ($p = 0.86$; no significant effect of genotype).

(B) Stick representation of hindlimb movement at the chronic stage during treadmill locomotion shown for spontaneous and daily agonist exposure groups (below: concurrent limb endpoint trajectories and velocity vector at swing onset, together with ipsilesional hindlimb oscillations; dark gray bars, stance; red bars, dragging).

(C) Histogram plots reporting mean values of scores on recovery-related PC1 (25% explained variance) performed on ipsilesional gait patterns (average of 10–20 gait cycles/mouse, 108 parameters per gait cycle; $n = 4$ [agonist exposure] or 9 [spontaneous] wild-type and $n = 5$ [both conditions] *Egr3* mutant mice). Within genotype, scores are not different between spontaneous and daily agonist exposure groups before injury and throughout recovery process.

(D) Factor loadings (correlation of kinematic parameter and recovery-associated PC) of PC1 for all parameters (individual dots) of the two conditions (spontaneous recovery, daily agonist exposure) were correlated against each other for wild-type and *Egr3* mutant mice. Strong positive correlation represents similar recovery process in both spontaneous and agonist exposure groups for both genotypes.

Error bars, SEM; ns, not significant; acute, 3 days postinjury; chronic, 7 weeks postinjury. See also Figures S1 and S2.

gigantocellular, spinal vestibular, and red nucleus, as well as in M1 motor cortex (Figures 5B, 5C, and S5B). These findings reveal an absence of significant baseline differences between wild-type and *Egr3* mutant mice, allowing direct comparison of descending projection neuron populations across genotypes after injury.

At the acute stage, the majority of ipsilesional brainstem nuclei were not labeled. This depletion results from the disrupted access of ipsilaterally projecting brainstem nuclei to circuits below lesion. Lesion also disconnected contralaterally projecting descending pathways that decussate above lesion, e.g., leading to a lack of Rab-FP^{ON} neurons in M1 motor cortex and the red nucleus (Figures 5B, 5C, and S5B). In contrast, we detected

a fraction of retrogradely marked spinal vestibular neurons residing in the ipsilesional brainstem (Figures 5B, 5C, and S5B). Axons of such neurons cross the midline above lesion, descend the spinal cord contralaterally, and establish collaterals crossing the midline a second time below lesion. We classified neurons with such axonal trajectories as dual midline-crossing projection neurons.

At chronic stages, we detected substantial reorganization of ipsilesional brainstem pathways in wild-type mice. Magnocellular, gigantocellular, and pontine nuclei contained significantly more ipsilesional Rab-FP^{ON} neurons than at acute stages (Figures 5B, 5C, and S5B). The presence of retrogradely labeled neurons in the ipsilesional brainstem thus implies that their axons

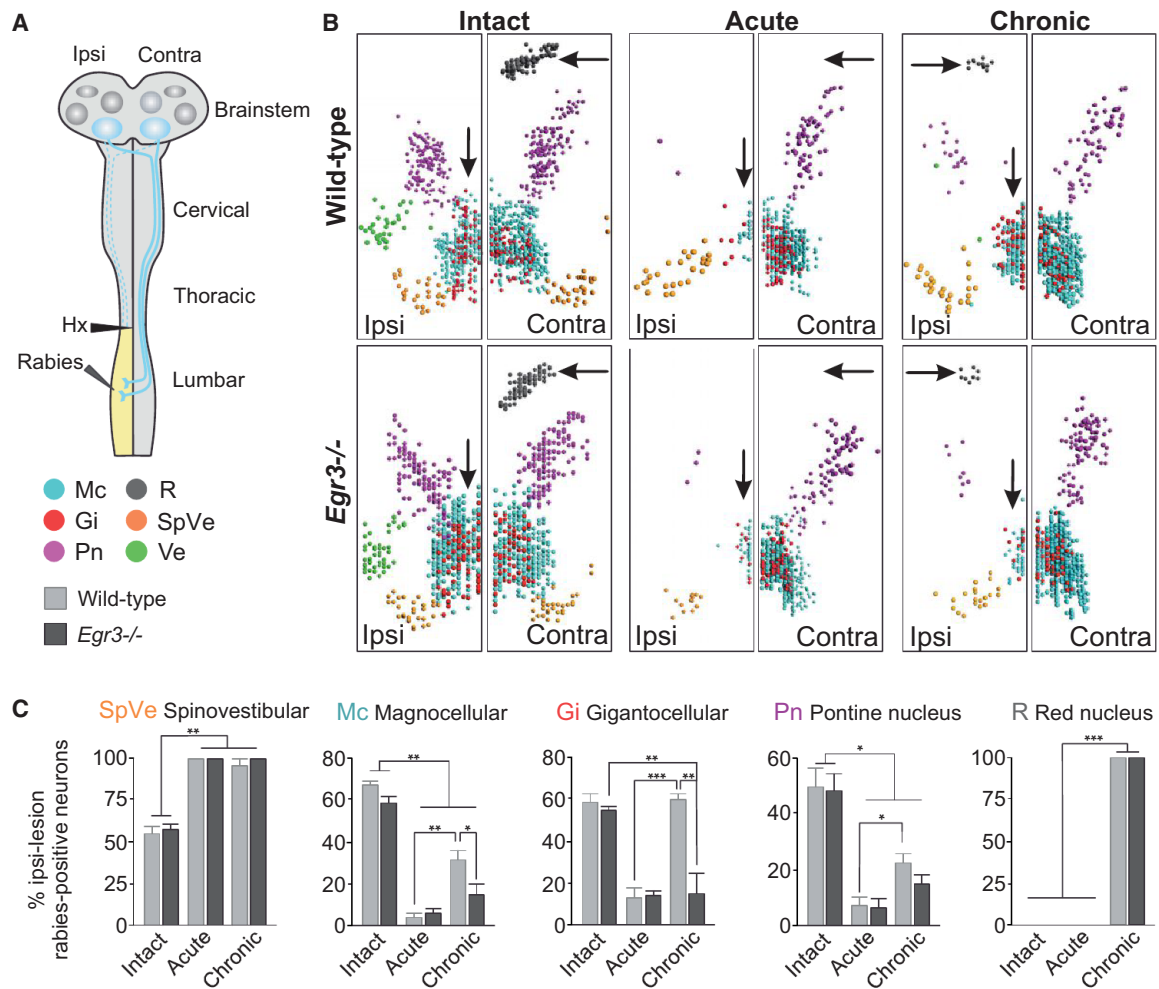


Figure 5. Reduced Injury Responses in Brainstem Pathways of *Egr3* Mutant Mice

(A) Diagram illustrating rabies virus injection strategy to retrogradely label brainstem neurons with descending projections to ipsilesional lumbar spinal cord (yellow). Bottom: display of different brainstem nuclei (Esposito et al., 2014; Paxinos and Franklin, 2012): Mc, magnocellular nucleus; Gi, Gigantocellular nucleus; Pn, Pontine nucleus; R, red nucleus; SpVe, spinal vestibular nucleus; Ve, vestibular nucleus.

(B) Top-down snapshots of 3D brainstem reconstructions in wild-type (top) and *Egr3* mutant (bottom) mice at intact, acute, and chronic stages (ipsi- and contralesional halves of the reconstruction displayed separately; each neuron represented by single dot; for color-code, see A).

(C) Quantification of brainstem reconstruction data ($n = 3$ each for intact and acute wild-type and *Egr3* mutant, $n = 4$ each for chronic wild-type and *Egr3* mutant) displaying percentage of ipsilesional neurons of entire rabies-marked respective subpopulation.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; error bars, SEM; acute, 3 days postinjury; chronic, 7 weeks postinjury; Hx, hemisection. See also Figure S5.

cross the midline twice to establish novel dual midline-crossing pathways. In contrast, *Egr3* mutant mice showed reduced levels of brainstem projection reorganization, with no significant differences in ipsilesional Rab-FP^{ON} neurons in magnocellular, gigantocellular, and pontine nuclei at chronic compared to acute stages (Figures 5B, 5C, and S5B). In contrast, we did not detect significant differences for neurons in the red nucleus between wild-type and *Egr3* mutant mice. Together, these results suggest that functional muscle spindle feedback facilitates rearrangement of specific descending pathways from the brainstem, but interestingly, not all populations were affected equally.

Next, we evaluated the effect of hemisection lesion on spinal projection neurons. We used rabies viruses to label neurons

through their axonal projections (Figures S6A–S6C). We also exploited a transsynaptic virus-based approach with monosynaptic restriction to capture synaptic connectivity (Wickersham et al., 2007) (Figure 6). Both approaches revealed similar distribution patterns of spinal projection neurons across multiple segments of the spinal cord at intact stages in both wild-type and *Egr3* mutant mice (Figures 6 and S6A–S6C). At acute stages, only few ipsilesional spinal projection neurons exhibited dual midline-crossing circuitry (Figures 6B and 6C; Figures S6B and S6C). In contrast, contralesional spinal neurons were abundantly marked by Rab-FP. At chronic stages, Rab-FP injections in wild-type mice revealed a prominent increase in the percentage of ipsilesional neurons compared to acute stages (Figures 6B and

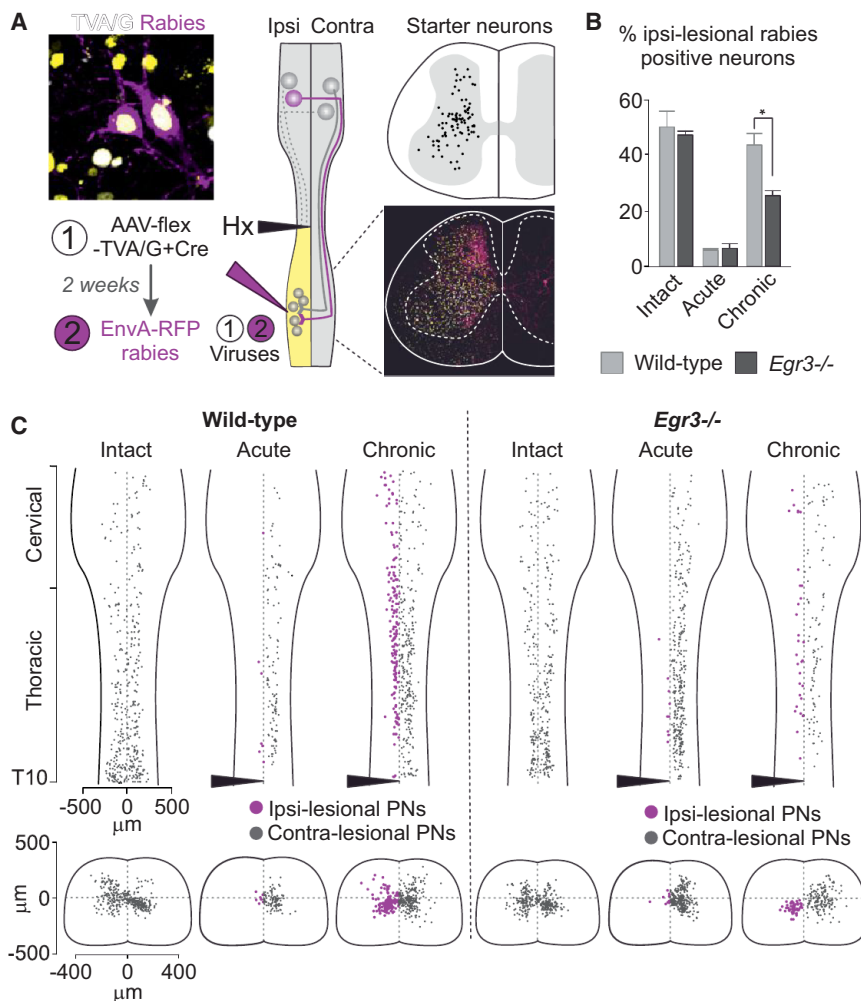


Figure 6. Detour Circuit Formation after Spinal Cord Injury Is Reduced in *Egr3* Mutant Mice

(A) Diagram illustrating monosynaptic rabies virus injection strategy to retrogradely mark descending spinal projection neurons with synaptic connections to ipsilesional neurons below lesion (injection at L2–L5; yellow; neurons with dotted axons, severed by injury; magenta, dual-crossing ipsilesional neurons). Top-left corner: example of triple-labeled (TVA/G/Rabies) neurons. Right: low-resolution view and reconstruction of triple-positive starter neurons of representative spinal cord section.

(B) Quantification of percentage of ipsilesional rabies positive spinal projection neurons above lesion with connections to ipsilesional starter neurons ($n = 3$ each for intact and acute wild-type and *Egr3* mutant; $n = 4$ for chronic wild-type; $n = 5$ for *Egr3* mutant).

(C) 3D reconstructions of supraspinal spinal projection neurons with connection to ipsilesional lumbar circuits below lesion (yellow) in wild-type (left) and *Egr3* mutant (right) mice at intact, acute, and chronic stages in top-down longitudinal view (top) and transverse section (below) view (filled triangle, lesion position; gray line, midline; magenta, ipsilesional neurons).

* $p < 0.05$; error bars, SEM; acute, 3 days post-injury; chronic, 7 weeks postinjury; Hx, hemisection. See also Figure S6.

6C; Figures S6B and S6C). In *Egr3* mutant mice, however, their percentage above lesion was significantly lower than in wild-type mice (Figures 6B and 6C; Figures S6B and S6C). In contrast, we did not detect distribution differences of spinal projection neurons between genotypes below lesion (Figures S6D and S6E). Together, these results demonstrate that *Egr3* mutant mice exhibit a deficiency in the establishment of dual-crossing detour circuits involving multiple populations of descending projection neurons, whereas these are abundantly detected in wild-type mice.

Spinal Projection Neurons Connect to Deprived Circuits by Distinct Mechanisms

To gain insight into the cellular mechanisms responsible for the emergence of ipsilesional dual-crossing detour circuits after hemisection, we devised anterograde virus-mediated tracing experiments from supraspinal spinal segments. We used injections of viruses that allow visualizing axons and synapses of specific subpopulations of descending projection neurons. This strategy also enabled evaluation of circuit reorganization from contralesional neurons that are less amenable to assess with retrograde tracing approaches due to partial persistence of projections to ipsilesional circuits after lesion.

We conducted a series of anterograde tracing experiments from ipsilesional or contralesional circuits above injury (Figure 7A) by unilateral coinjections of AAV-Tomato and AAV-SynGFP or AAV-SynMyc to anterogradely track axonal trajectories and synaptic arborizations (Figures 7 and S7). In intact mice of both genotypes, cervical (C5–7) and thoracic (T7–8) neurons projected to lumbar levels bilaterally (Figure S7A). After injury, ipsilesional injections target commissural neurons with axonal tracts descending contralateral to cell body position, whereas contralesional injections target ipsilateral projection neurons with axonal tracts ipsilateral to cell body position (Figure 7A).

We found that for both ipsi- and contralesional injections, descending axon tracts were present bilaterally in the white matter above lesion (Figure S7B). After hemisection, tracts only persisted on the contralesional side below lesion, demonstrating that axon collaterals at lumbar levels were exclusively derived from neurons projecting through contralesional tracts (Figures 7A and S7B). Next, we determined the frequency of midline-crossing axon collaterals below lesion at 2 weeks postlesion, the earliest possible tracing time point, and at chronic stages (Figures 7B and 7C). At 2 weeks postinjury, no difference was observed for ipsi- or contralesional populations and their midline-crossing frequency between wild-type and *Egr3* mutant mice (Figure 7C). At chronic stages, *Egr3* mutants exhibited a significantly reduced number of midline-crossing axons derived from ipsilesional populations compared to wild-type

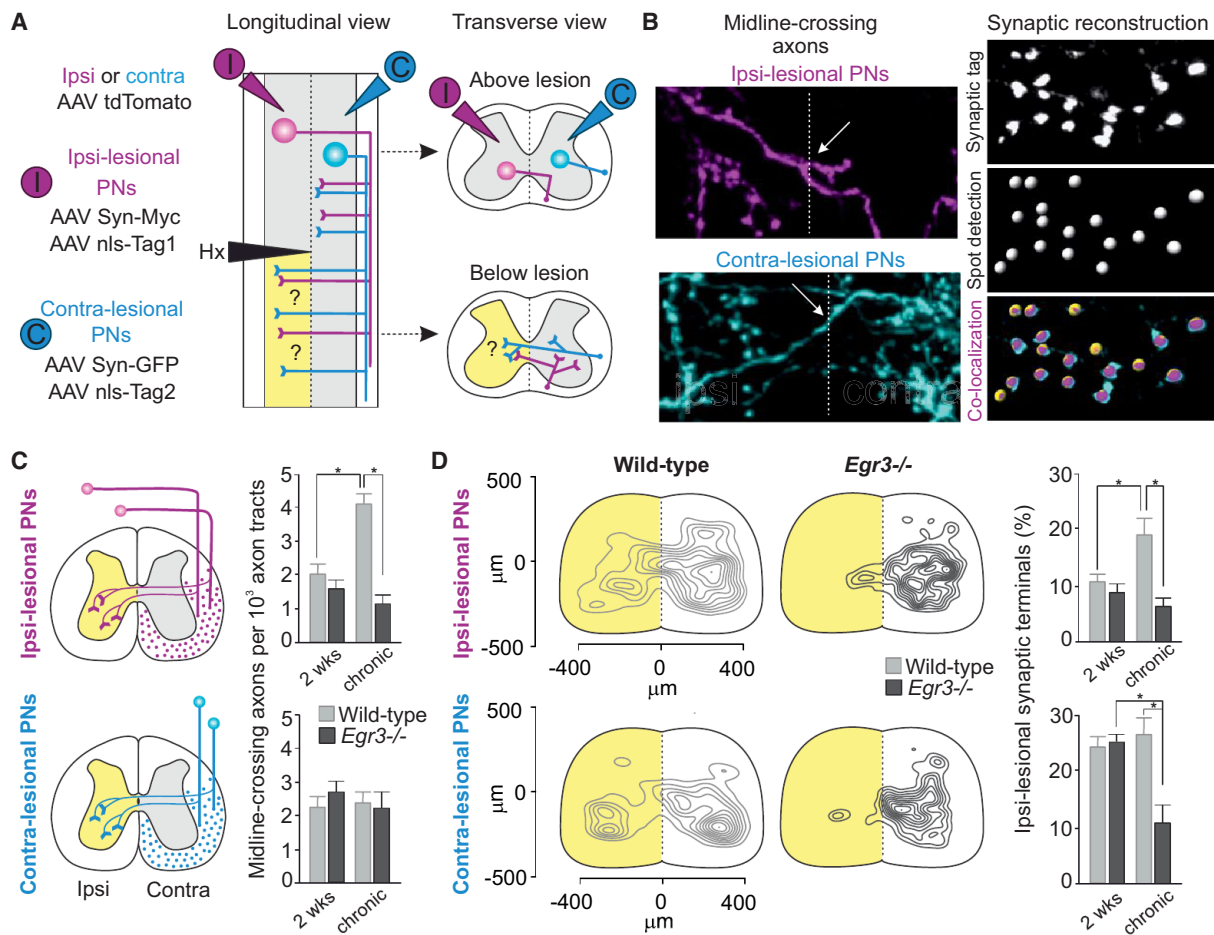


Figure 7. Distinct Reconnection Mechanisms for Spinal Projection Neuron Subpopulations

(A) Diagrams illustrating intraspinal injection scheme to anterogradely visualize axons (tdTomato) and synapses (synaptically tagged proteins) of ipsilesional (magenta) or contralesional (blue) spinal projection neurons residing above lesion (nuclear markers confirm unilaterality of injection). Top-down longitudinal and cross-section projected views shown (yellow, ipsilesional territory below lesion).

(B) Examples of midline-crossing axons and synaptic terminal analysis with high-resolution imaging and Imaris spot detection.

(C) Frequency analysis of midline-crossing axons originating from ipsi- and contralesional cervical spinal projection neurons, normalized to number of marked axons in contralesional white matter tracts below lesion (ipsilesion: $n = 3$ each for wild-type and *Egr3* mutant for both time points; contralesion: $n = 4$ for wild-type and $n = 3$ *Egr3* mutant for 2 week time point; $n = 4$ for wild-type and $n = 5$ *Egr3* mutant for chronic analysis).

(D) Quantitative analysis of distribution and density of synaptic terminals in the spinal cord below injury, originating from ipsi- and contralesional cervical spinal projection neurons (yellow, ipsilesional territory below lesion; ipsilesion, $n = 4$ for wild-type and $n = 5$ for *Egr3* mutant for 2 week time point; $n = 5$ for wild-type and $n = 6$ for *Egr3* mutant for chronic analysis; contra-lesion, $n = 4$ for wild-type and $n = 6$ for *Egr3* mutant for 2 week time point; $n = 5$ each for wild-type and *Egr3* mutant for chronic analysis). Contour plots show overall distribution of terminals from one chronic animal for each genotype; histogram plots display percentage of ipsilesional synaptic terminals at analyzed segmental spinal levels.

* $p < 0.05$; error bars, SEM; Hx, hemisection; PN, projection neuron; chronic, 7 weeks postinjury. See also Figure S7.

mice, whereas no significant difference was observed for contralesional populations (Figures 7C and S7F). Together, these findings indicate that the absence of muscle spindle feedback impairs the ability to establish de novo dual midline-crossing axons originating from ipsilesional spinal projection neurons and that these anatomical differences between the two genotypes become apparent later than 2 weeks after injury.

Next, we quantified ipsilesional synaptic arborization of midline-crossing axons (Figures 7B and 7D; Figures S7D and S7G). We reconstructed synaptic puncta at high resolution, yielding quantitative information on the spatial distribution and

number of synaptic terminals (Figure 7B). Analysis of ipsi- and contralesional projection neurons revealed comparable synaptic innervation above lesion between wild-type and *Egr3* mutant mice (Figure S7D). In wild-type mice at chronic stages below lesion, synaptic input to ipsilesional gray matter targeted the ventrolateral quadrant, which contains many locomotor interneurons and motor neurons (Figures 7D and S7G). In contrast, the distribution of synaptic input beyond the midline in *Egr3* mutant mice was primarily confined to medially located territory (Figures 7D and S7G). The observed increase in synaptic terminals in wild-type mice was not present 2 weeks postlesion,

in agreement with the corresponding time course of midline-crossing axon elaboration (Figures 7C and 7D). For contralateral projection neurons, we also detected lower synaptic terminal density in the ipsilesional gray matter below lesion in *Egr3* mutant mice at chronic stages compared to wild-type mice, despite a similar number of midline-crossing axons in both genotypes (Figures 7C and 7D; Figures S7D and S7G). Strikingly, however, we found a selective decrease in the density of synaptic terminals between 2 weeks postlesion and chronic stages in *Egr3* mutant mice, ultimately leading to the observed lower terminal density compared to wild-type mice (Figure 7D). Together, our findings provide evidence that after lateral hemisection spinal cord injury, muscle spindle feedback enhances the process of axonal and synaptic rearrangements of multiple descending spinal projection neuron populations through distinct mechanisms.

DISCUSSION

Spinal cord injuries lead to immediate motor dysfunction because of separation of descending control pathways from local spinal circuits. Various degrees of functional recovery occur after incomplete injury. However, the likely involvement of numerous circuit elements paired with the limited understanding of their precise organization and function within the hierarchy of motor control pathways have posed challenges for gaining mechanistic insight in the process of functional recovery. Here, we demonstrate that muscle spindle feedback circuits are essential to direct locomotor recovery after lateral hemisection spinal cord injury and that the lack of this specific sensory channel affects the ability of descending projection neurons to undergo efficient circuit reorganization after injury. We discuss our findings with an emphasis on the role of sensory feedback circuits in locomotor improvement after injury and the mechanisms by which circuit rearrangements parallel and influence the recovery process.

Task-Specific Locomotor Recovery after Spinal Cord Injury in Wild-Type Mice

Wild-type mice improve basic locomotor function after hemisection spinal cord injury to a significant extent. In contrast, they remain severely compromised in their ability to carry out precision ladder walking. These findings underscore the need for task-specific communication channels between supraspinal and spinal circuits, some of which do not recover after injury. A possible model to explain these findings is that upon establishment of multistep synaptic relays, a comparatively crude wiring of descending circuit elements is sufficient to drive disconnected ipsilesional spinal circuits below lesion for regaining basic locomotor function. Newly established descending connections can interact with an already wired repertoire of local spinal circuits able to coordinate basic locomotor behaviors. In contrast, precision tasks likely require specific and refined descending circuit connectivity. In addition, complex tasks may depend more heavily on information conveyed through ascending pathways, which exhibit enduring dysfunction after spinal cord injury (Kaas et al., 2008; Martinez et al., 2010). Taken together, these observations suggest that distinct neuronal circuit ele-

ments are responsible and necessary for the re-establishment of task-specific functions.

Role of Muscle Spindle Feedback Circuits in Locomotor Recovery after Spinal Cord Injury

Muscle spindle afferents constitute a minor fraction of DRG sensory neurons (Scott, 1992), but our results demonstrate that they are essential to promote locomotor recovery after incomplete spinal cord lesion. Why does deprivation of a specific sensory channel lead to such profound impairment? Each class of functionally distinct sensory neurons exhibits lamina-specific axonal terminations in the spinal cord (Brown, 1981). While cutaneous and mechanoreceptive afferents target dorsal horn neurons, proprioceptive afferents terminate more ventrally, raising the possibility that these differential synaptic connectivity profiles may contribute to their role in the recovery process.

A primary mode of action by muscle spindle afferents in facilitating recovery may involve recruitment of motor circuits through their unique connections. Targeted circuit elements include motor neurons and core components of ventral locomotor interneuron circuits that have recently been demonstrated to play important roles in the regulation of extensor-flexor alternation (Talpalari et al., 2011; Zhang et al., 2014) and rhythm generation (Dougherty et al., 2013) in the mouse. The pivotal role of muscle spindle feedback in promoting locomotor improvement after lateral hemisection observed here might therefore be at least in part attributed to their direct synaptic access to these neurons. Specifically, muscle spindle afferents are embedded in a highly selective central synaptic connectivity matrix. Transfer of muscle-specific information to functionally distinct interneurons that directly activate motor neurons or mediate reciprocal inhibition between motor neurons is a key feature of these neuronal networks (Jankowska and Edgley, 1993; McCreary and Rybak, 2008; Pearson, 2008; Wang et al., 2008; Windhorst, 2007). Muscle spindle afferent recruitment after injury may strengthen these specific spinal circuits and their connections (Petruska et al., 2007), whereas their functional absence in *Egr3* mutants might contribute to the severe impairment in recovery.

An alternative or complementary possibility is that muscle spindle afferents release factors in an activity-dependent manner, which in turn promote circuit reorganization in the spinal cord. For instance, retrograde trophic support by the neurotrophin NT-3 strengthens synaptic connections (Boyce and Mendell, 2014; Chen et al., 2002; Oakley et al., 1997). Moreover, the amount of physical activity influences baseline BDNF expression in the spinal cord after traumatic injury (Ying et al., 2008), an effect that may be mediated by recruitment of muscle spindle feedback circuits. In our experiments, we found no difference in the degree of spontaneous cage activity between wild-type and *Egr3* mutants after hemisection spinal cord injury, excluding disparity in physical activity as a possible reason for differential recovery. In addition, daily application of monoaminergic agents to enhance activity of local spinal circuits in an attempt to bypass reduced sensory feedback in *Egr3* mutant mice was inefficient in overcoming the severely limited recovery in *Egr3* mutant mice. These findings demonstrate that muscle spindle afferents, despite being a numerically minor sensory

neuron population, play an instrumental and selective role in promoting functional recovery after spinal cord injury.

Formation of Spinal Detour Circuits Parallels Locomotor Recovery

Regaining locomotor function of the ipsilesional hindlimb after thoracic hemisection requires the establishment of detour circuits that reconnect descending pathways to deprived locomotor circuits below lesion. The formation of such detour circuits to functionally bridge the injury site depends on local axon growth and reorganization of synaptic connectivity within existing descending circuit modules (Ballermann and Fouad, 2006; Bareyre et al., 2004; Courtine et al., 2008; Jankowska and Edgley, 2006; Rosenzweig et al., 2010; van den Brand et al., 2012). We demonstrate that in wild-type mice, injury-induced circuit-level responses involve the deployment of specific patterns of axonal growth and synaptic arborization from distinct populations of supraspinal and spinal projection neurons.

Our anatomical mapping to identify injury-responsive descending circuit elements above lesion demonstrates that reduced compensatory responses to injury are widespread in *Egr3* mutants. These alterations include ipsi- and contralesional spinal projection neurons at multiple spinal segments and specific descending pathways from the brainstem. Perturbation or silencing of any identified specific neuron population alone in wild-type mice is therefore unlikely to recapitulate the dramatic lack of recovery observed in *Egr3* mutant mice. On the other hand, experimental attempts to specifically target a majority of neurons undergoing novel collateral formation after injury would require injections at multiple central nervous system (CNS) sites, likely themselves inducing behavioral repercussions. In addition, even if successful, such approaches would interfere with the function of targeted neurons in their entirety, and not just with the newly formed collaterals.

How does muscle spindle feedback facilitate de novo circuit formation? While we cannot rule out multifaceted circuit-level effects influencing the recovery process, we favor a model in which muscle spindle feedback circuits act primarily on ipsilesional circuits below the injury site to promote the formation of compensatory connections to deprived circuits. In agreement with such a model, identified brainstem populations do not receive direct synaptic input from muscle spindle afferents, implying that at least for these populations such input is not essential to trigger circuit reorganization. Mechanistically, the assembly of novel circuits in the adult nervous system may be achieved through stabilization of nascent axon collaterals involving Hebbian plasticity reinforced by muscle spindle afferent input. Growth and stabilization of axons in the developing nervous system suggests that such mechanisms act in highly cell-type-specific patterns (Andreae and Burrone, 2014).

To gain insight into how defined neuronal populations respond to injury, we focused our anterograde synaptic mapping analysis on spinal projections neurons. Comparison of wild-type and *Egr3* mutant mice uncovered distinct responses for specific spinal populations. Ipsilesional descending spinal projection neurons in *Egr3* mutants exhibited both a reduction in dual midline-crossing axons and decreased ipsilesional synaptic arborization below lesion, occurring later than 2 weeks

after injury. In contrast, contralesional counterparts only showed restricted arborization of synaptic terminals without disruption in midline-crossing axons. These synaptic differences however can be attributed to synaptic pruning in the absence of muscle spindle feedback rather than additional synaptic growth in wild-type mice. These findings also suggest that the majority of injury-responsive contralesional spinal projection neurons already possess midline-crossing collaterals at intact stages, providing an explanation for why this parameter is not affected in *Egr3* mutants compared to wild-type mice at chronic stages.

In summary, our study demonstrates that one specific sensory channel has an executive role in directing restoration of hindlimb motor function and facilitating multifaceted circuit reorganization after incomplete spinal cord injury. These findings stress the importance of exploiting muscle spindle feedback circuits in the design of rehabilitative strategies after spinal cord injury. Epidural stimulation of lumbar segments facilitates motor control and leads to improved functional recovery in animal models and paraplegic individuals (Angeli et al., 2014; van den Brand et al., 2012). This treatment paradigm may at least in part act through the recruitment of myelinated sensory feedback circuits (Capogrosso et al., 2013). Refined experimental strategies to specifically modulate muscle spindle feedback channels open innovative therapeutic avenues to pursue in the future. Similar concepts may apply to other traumatic CNS disorders, such as stroke or brain injury, which heavily rely on plasticity of both supraspinal and spinal descending pathways to regain functional capacities after lesion.

EXPERIMENTAL PROCEDURES

Mouse Genetics and Surgeries

Mice used were from a local colony containing the *Egr3* mutant allele previously described (Tourtellotte and Milbrandt, 1998). Surgical procedures for hemisection injury and EMG implantation have been described previously (Courtine et al., 2008) and were performed under full general anesthesia with isoflurane in oxygen-enriched air (1%–2%). Local Swiss veterinary offices approved all the procedures. Details on mice and surgical procedures are described in the Extended Experimental Procedures.

Behavioral Analysis

Whole-body kinematics were recorded using the high-speed motion capture system Vicon (Vicon Motion Systems), combining 10–12 infrared cameras (200 Hz) (van den Brand et al., 2012). Parameters describing kinematic and EMG characteristics were computed using custom-written MATLAB scripts (van den Brand et al., 2012). Behavioral tests included overground locomotion on an elevated runway, stepping on a motorized treadmill (Robomedica), elevated horizontal ladder, and swimming. To quantify task- and genotype-specific gait characteristics prior to injury and throughout the recovery process after hemisection spinal cord injury, we implemented a multistep statistical procedure based on PC analysis (Dominici et al., 2012). A flowchart explaining the various steps of this analysis can be found in Figure S2. For behavioral monitoring of home cage activity, spontaneous activity was surveyed for each mouse during 20 min. Additional information on recordings, postprocessing and behavioral tasks are available in the Extended Experimental Procedures and Figure S2.

Anatomical Tracing Experiments

Rabies viruses and AAVs were amplified and purified from local viral stocks following established protocols (Esposito et al., 2014; Pivetta et al., 2014; Wickersham et al., 2010). Additional information on production and injection

of viruses, antibodies, imaging, and anatomical quantification can be found in the Extended Experimental Procedures.

Statistical Analysis

All data are reported as mean values \pm SEM. All statistical evaluations were performed using GraphPad Prism (v. 6.0) (Prism, GraphPad Software) using unpaired Student's *t* test (Figures 1C, 1D, 2D, 5C, 6B, 7C, and 7D; Figures S3A, S3B, S4D, S6C, S7A, S7B, S7D, S7F, and S7G), two-way ANOVA for repeated measurement (Figures 1D, 3D, 3E, 4A, and 4C; Figures S3D, S3E, S5A, and S5B), and one-way ANOVA for repeated measurements (Figure S5C), followed by post hoc comparisons (Šidák-Bonferroni). The significance level for behavioral analysis was set as $|R \text{ value}| > 0.5$ and $p < 0.05$, respectively. Significance level is defined as follows for all analyses performed: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and five movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.019>.

AUTHOR CONTRIBUTIONS

A.T., G.C., and S.A. initiated the project. A.T. and I.V. performed behavioral experiments and analysis; A.T. performed anatomy experiments and analysis. All authors were involved in interpretation of experiments and contributed to writing the paper.

ACKNOWLEDGMENTS

We are grateful to M. Sigrist, M. Mielich, P. Capelli, and S. Esposito for expert technical help, M. Kirschmann, S. Bourke, and L. Gelman from the FMI imaging facility, N. Ehrenfurchter from the Biozentrum Imaging facility for help and advice with image acquisition and analysis, and P. Caroni for discussions and comments on the manuscript. A.T. and S.A. were supported by an ERC Advanced Grant, the Swiss National Science Foundation, the Kanton Basel-Stadt, and the Novartis Research Foundation. I.V. was supported by the European Neuroscience Campus (ENC) network. G.C. was supported by an ERC Starting Grant and the International Foundation for Research in Paraplegia (IRP).

Received: August 13, 2014

Revised: November 5, 2014

Accepted: November 11, 2014

Published: December 18, 2014

REFERENCES

- Andreae, L.C., and Burrone, J. (2014). The role of neuronal activity and transmitter release on synapse formation. *Curr. Opin. Neurobiol.* 27, 47–52.
- Angeli, C.A., Edgerton, V.R., Gerasimenko, Y.P., and Harkema, S.J. (2014). Altering spinal cord excitability enables voluntary movements after chronic complete paralysis in humans. *Brain* 137, 1394–1409.
- Arshavskii, I.I., Kots, I.A., Orlovskii, G.N., Rodionov, I.M., and Shik, M.L. (1965). [Study of biomechanics of running dogs]. *Biofizika* 10, 665–671.
- Ballermann, M., and Fouad, K. (2006). Spontaneous locomotor recovery in spinal cord injured rats is accompanied by anatomical plasticity of reticulospinal fibers. *Eur. J. Neurosci.* 23, 1988–1996.
- Bareyre, F.M., Kerschensteiner, M., Raineteau, O., Mettenleiter, T.C., Weinmann, O., and Schwab, M.E. (2004). The injured spinal cord spontaneously forms a new intraspinal circuit in adult rats. *Nat. Neurosci.* 7, 269–277.
- Bouyer, L.J., and Rossignol, S. (2003). Contribution of cutaneous inputs from the hindpaw to the control of locomotion. II. Spinal cats. *J. Neurophysiol.* 90, 3640–3653.
- Boyce, V.S., and Mendell, L.M. (2014). Neurotrophins and spinal circuit function. *Front. Neural Circuits* 8, 59.
- Brown, A.G. (1981). *Organization of the Spinal Cord* (New York: Springer Verlag).
- Capogrosso, M., Wenger, N., Raspopovic, S., Musienko, P., Beauparlant, J., Bassi Luciani, L., Courtine, G., and Micera, S. (2013). A computational model for epidural electrical stimulation of spinal sensorimotor circuits. *J. Neurosci.* 33, 19326–19340.
- Chen, H.H., Tourtellotte, W.G., and Frank, E. (2002). Muscle spindle-derived neurotrophin 3 regulates synaptic connectivity between muscle sensory and motor neurons. *J. Neurosci.* 22, 3512–3519.
- Courtine, G., Song, B., Roy, R.R., Zhong, H., Herrmann, J.E., Ao, Y., Qi, J., Edgerton, V.R., and Sofroniew, M.V. (2008). Recovery of supraspinal control of stepping via indirect propriospinal relay connections after spinal cord injury. *Nat. Med.* 14, 69–74.
- Curt, A., Van Hedel, H.J., Klaus, D., and Dietz, V.; EM-SCI Study Group (2008). Recovery from a spinal cord injury: significance of compensation, neural plasticity, and repair. *J. Neurotrauma* 25, 677–685.
- Dietz, V., and Fouad, K. (2014). Restoration of sensorimotor functions after spinal cord injury. *Brain* 137, 654–667.
- Dougherty, K.J., Zagoraiou, L., Satoh, D., Rozani, I., Doobar, S., Arber, S., Jessell, T.M., and Kiehn, O. (2013). Locomotor rhythm generation linked to the output of spinal shox2 excitatory interneurons. *Neuron* 80, 920–933.
- Eccles, J.C., Eccles, R.M., and Lundberg, A. (1957). The convergence of monosynaptic excitatory afferents on to many different species of alpha motoneurons. *J. Physiol.* 137, 22–50.
- Edgerton, V.R., Courtine, G., Gerasimenko, Y.P., Lavrov, I., Ichiyama, R.M., Fong, A.J., Cai, L.L., Otsu, C.K., Tillakaratne, N.J., Burdick, J.W., and Roy, R.R. (2008). Training locomotor networks. *Brain Res. Brain Res. Rev.* 57, 241–254.
- Esposito, M.S., Capelli, P., and Arber, S. (2014). Brainstem nucleus MdV mediates skilled forelimb motor tasks. *Nature* 508, 351–356.
- Gruner, J.A., and Altman, J. (1980). Swimming in the rat: analysis of locomotor performance in comparison to stepping. *Exp. Brain Res.* 40, 374–382.
- Halbertsma, J.M. (1983). The stride cycle of the cat: the modelling of locomotion by computerized analysis of automatic recordings. *Acta Physiol. Scand. Suppl.* 521, 1–75.
- Jankowska, E., and Edgley, S. (1993). Interactions between pathways controlling posture and gait at the level of spinal interneurons in the cat. *Prog. Brain Res.* 97, 161–171.
- Jankowska, E., and Edgley, S.A. (2006). How can corticospinal tract neurons contribute to ipsilateral movements? A question with implications for recovery of motor functions. *Neuroscientist* 12, 67–79.
- Kaas, J.H., Qi, H.X., Burish, M.J., Gharbawie, O.A., Onifer, S.M., and Massey, J.M. (2008). Cortical and subcortical plasticity in the brains of humans, primates, and rats after damage to sensory afferents in the dorsal columns of the spinal cord. *Exp. Neurol.* 209, 407–416.
- Knikou, M., and Mummidisetti, C.K. (2014). Locomotor training improves premotoneuronal control after chronic spinal cord injury. *J. Neurophysiol.* 111, 2264–2275.
- Lavrov, I., Courtine, G., Dy, C.J., van den Brand, R., Fong, A.J., Gerasimenko, Y., Zhong, H., Roy, R.R., and Edgerton, V.R. (2008). Facilitation of stepping with epidural stimulation in spinal rats: role of sensory input. *J. Neurosci.* 28, 7774–7780.
- Maier, I.C., and Schwab, M.E. (2006). Sprouting, regeneration and circuit formation in the injured spinal cord: factors and activity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1611–1634.
- Martinez, M., Delcourt, M., Russier, M., Zennou-Azogui, Y., Xerri, C., Coq, J.O., and Brezun, J.M. (2010). Differential tactile and motor recovery and cortical map alteration after C4–C5 spinal hemisection. *Exp. Neurol.* 221, 186–197.
- McCreary, D.A., and Rybak, I.A. (2008). Organization of mammalian locomotor rhythm and pattern generation. *Brain Res. Brain Res. Rev.* 57, 134–146.
- Oakley, R.A., Lefcort, F.B., Clary, D.O., Reichardt, L.F., Prevett, D., Oppenheim, R.W., and Frank, E. (1997). Neurotrophin-3 promotes the differentiation

- of muscle spindle afferents in the absence of peripheral targets. *J. Neurosci.* 17, 4262–4274.
- Paxinos, G., and Franklin, K.B. (2012). *The Mouse Brain in Stereotaxic Coordinates*, Fourth Edition (San Diego: Elsevier).
- Pearson, K.G. (2008). Role of sensory feedback in the control of stance duration in walking cats. *Brain Res. Brain Res. Rev.* 57, 222–227.
- Petruska, J.C., Ichiyama, R.M., Jindrich, D.L., Crown, E.D., Tansey, K.E., Roy, R.R., Edgerton, V.R., and Mendell, L.M. (2007). Changes in motoneuron properties and synaptic inputs related to step training after spinal cord transection in rats. *J. Neurosci.* 27, 4460–4471.
- Pivetta, C., Esposito, M.S., Sigrist, M., and Arber, S. (2014). Motor-circuit communication matrix from spinal cord to brainstem neurons revealed by developmental origin. *Cell* 156, 537–548.
- Ramon y Cajal, S. (1928). *Degeneration and Regeneration of the Nervous System* (Oxford: Oxford University Press).
- Rosenzweig, E.S., Courtine, G., Jindrich, D.L., Brock, J.H., Ferguson, A.R., Strand, S.C., Nout, Y.S., Roy, R.R., Miller, D.M., Beattie, M.S., et al. (2010). Extensive spontaneous plasticity of corticospinal projections after primate spinal cord injury. *Nat. Neurosci.* 13, 1505–1510.
- Rossignol, S., and Frigon, A. (2011). Recovery of locomotion after spinal cord injury: some facts and mechanisms. *Annu. Rev. Neurosci.* 34, 413–440.
- Rossignol, S., Dubuc, R., and Gossard, J.P. (2006). Dynamic sensorimotor interactions in locomotion. *Physiol. Rev.* 86, 89–154.
- Roy, R.R., Harkema, S.J., and Edgerton, V.R. (2012). Basic concepts of activity-based interventions for improved recovery of motor function after spinal cord injury. *Arch. Phys. Med. Rehabil.* 93, 1487–1497.
- Scott, S.A. (1992). *Sensory Neurons: Diversity, Development and Plasticity* (New York: Oxford University Press).
- Talpalar, A.E., Endo, T., Löw, P., Borgius, L., Hägglund, M., Dougherty, K.J., Ryge, J., Hnasko, T.S., and Kiehn, O. (2011). Identification of minimal neuronal networks involved in flexor-extensor alternation in the mammalian spinal cord. *Neuron* 71, 1071–1084.
- Tello, F. (1907). La regeneration dans les voies optiques. *Trab. Lab. Invest. Biol. Univ. Madr.* 5, 237–248.
- Tourtellotte, W.G., and Milbrandt, J. (1998). Sensory ataxia and muscle spindle agenesis in mice lacking the transcription factor *Egr3*. *Nat. Genet.* 20, 87–91.
- van den Brand, R., Heutschi, J., Barraud, Q., DiGiovanna, J., Bartholdi, K., Huerlimann, M., Friedli, L., Vollenweider, I., Morad, E.M., Duis, S., et al. (2012). Restoring voluntary control of locomotion after paralyzing spinal cord injury. *Science* 336, 1182–1185.
- Wang, Z., Li, L., Goulding, M., and Frank, E. (2008). Early postnatal development of reciprocal Ia inhibition in the murine spinal cord. *J. Neurophysiol.* 100, 185–196.
- Wickersham, I.R., Sullivan, H.A., and Seung, H.S. (2010). Production of glycoprotein-deleted rabies viruses for monosynaptic tracing and high-level gene expression in neurons. *Nat. Protoc.* 5, 595–606.
- Wickersham, I.R., Lyon, D.C., Barnard, R.J., Mori, T., Finke, S., Conzelmann, K.K., Young, J.A., and Callaway, E.M. (2007). Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* 53, 639–647.
- Windhorst, U. (2007). Muscle proprioceptive feedback and spinal networks. *Brain Res. Bull.* 73, 155–202.
- Ying, Z., Roy, R.R., Zhong, H., Zdunowski, S., Edgerton, V.R., and Gomez-Pinilla, F. (2008). BDNF-exercise interactions in the recovery of symmetrical stepping after a cervical hemisection in rats. *Neuroscience* 155, 1070–1078.
- Zhang, J., Lanuza, G.M., Britz, O., Wang, Z., Siembab, V.C., Zhang, Y., Velasquez, T., Alvarez, F.J., Frank, E., and Goulding, M. (2014). V1 and v2b interneurons secure the alternating flexor-extensor motor activity mice require for limbed locomotion. *Neuron* 82, 138–150.
- Zörner, B., Filli, L., Starkey, M.L., Gonzenbach, R., Kasper, H., Röthlisberger, M., Bolliger, M., and Schwab, M.E. (2010). Profiling locomotor recovery: comprehensive quantification of impairments after CNS damage in rodents. *Nat. Methods* 7, 701–708.
- Zörner, B., Bachmann, L.C., Filli, L., Kapitza, S., Gullo, M., Bolliger, M., Starkey, M.L., Röthlisberger, M., Gonzenbach, R.R., and Schwab, M.E. (2014). Chasing central nervous system plasticity: the brainstem's contribution to locomotor recovery in rats with spinal cord injury. *Brain* 137, 1716–1732.

The Cellular and Molecular Basis of Direction Selectivity of A δ -LTMRs

Michael Rutlin,^{2,5,6} Cheng-Ying Ho,^{2,3,5,7} Victoria E. Abraira,^{1,2,5} Colleen Cassidy,⁴ C. Jeffery Woodbury,^{4,*} and David D. Ginty^{1,2,*}

¹Department of Neurobiology, Howard Hughes Medical Institute, Harvard Medical School, 220 Longwood Avenue, Boston, MA 02115, USA

²Solomon H. Snyder Department of Neuroscience, Howard Hughes Medical Institute

³Department of Pathology

The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205, USA

⁴Department of Zoology and Physiology, University of Wyoming, Laramie, WY 82071, USA

⁵Co-first author

⁶Present address: Department of Biochemistry and Molecular Biophysics, Howard Hughes Medical Institute, Columbia College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA

⁷Present address: Department of Pathology, Children's National Medical Center, 111 Michigan Avenue NW, Washington, DC 20010, USA

*Correspondence: woodbury@uwyo.edu (C.J.W.), david_ginty@hms.harvard.edu (D.D.G.)

<http://dx.doi.org/10.1016/j.cell.2014.11.038>

SUMMARY

The perception of touch, including the direction of stimulus movement across the skin, begins with activation of low-threshold mechanosensory neurons (LTMRs) that innervate the skin. Here, we show that murine A δ -LTMRs are preferentially tuned to deflection of body hairs in the caudal-to-rostral direction. This tuning property is explained by the finding that A δ -LTMR lanceolate endings around hair follicles are polarized; they are concentrated on the caudal (downward) side of each hair follicle. The neurotrophic factor BDNF is synthesized in epithelial cells on the caudal, but not rostral, side of hair follicles, in close proximity to A δ -LTMR lanceolate endings, which express TrkB. Moreover, ablation of BDNF in hair follicle epithelial cells disrupts polarization of A δ -LTMR lanceolate endings and results in randomization of A δ -LTMR responses to hair deflection. Thus, BDNF-TrkB signaling directs polarization of A δ -LTMR lanceolate endings, which underlies direction-selective responsiveness of A δ -LTMRs to hair deflection.

INTRODUCTION

Our ability to detect the direction of movement of stimuli in our sensory world is critical to survival; therefore, it is no surprise that a large portion of our sensory systems is devoted to the perception of stimulus movement across our environmental landscape. In the visual system, direction-selective retinal ganglion cells (DS-RGCs) and higher order visual centers, such as the visual area middle temporal (MT), are concerned with image movement across visual space (Wei and Feller, 2011). In the auditory system, the principal nuclei of the superior olivary complex process interaural time differences, which are critical for sound localization (Grothe et al., 2010). While the cells and cir-

cuits underlying detection and processing of visual and auditory direction-selective stimuli are becoming understood, little is known about how the direction of movement of stimuli acting on the skin, which is our largest sensory organ, is detected and processed.

The sense of touch allows us to recognize and manipulate objects held in our hands, detect innocuous or potentially harmful stimuli acting upon our bodies, and it enables physical communication for social bonding, sexual pleasure, and procreation. The neurobiological steps leading to the perception of touch begin with activation of low-threshold mechanoreceptors (LTMRs) by physical stimuli acting on the skin. LTMR cell bodies reside within dorsal root ganglia (DRG) and trigeminal ganglia and have one axonal branch that extends to the periphery and associates with a cutaneous mechanosensory end organ and another branch that penetrates the spinal cord and forms synapses upon second order neurons in the spinal cord dorsal horn and, in some cases, the dorsal column nuclei of the brainstem. LTMRs are sensitive to innocuous indentation, stroking, vibration, or stretch of the skin, and the deflection of hair follicles. Current challenges include defining mechanisms of unique tuning properties and functions of LTMR subtypes and determining how ensembles of LTMR activities are represented, integrated and processed in the CNS to give rise to the perception of touch.

Adrian and Zotterman (1926) first described the electrophysiological properties of sensory neurons that respond to hairy skin stimulation and their work laid the foundation and subsequent classification of the main LTMR types that associate with mammalian hairy skin (Zotterman, 1939). A β RA-LTMRs, field receptors (F-LTMRs), A β SAI-LTMRs, down (D-) hair follicle afferents/A δ -LTMRs, and C-LTMRs were initially defined based on stimulus response characteristics, the conduction velocity of their action potentials, adaptation properties, and the morphology of hairs with which they associate (Brown and Iggo, 1967; Burgess et al., 1968; Zotterman, 1939). A β RA-LTMRs and A β SAI-LTMRs have large myelinated axons, fast conduction velocities, and adapt rapidly or slowly, respectively, during sustained mechanical stimulation of the skin. While A β RA-LTMR subtypes are velocity detectors that respond to

skin indentation, movement of stimuli across the skin, and deflection of hair follicles, A β SAI-LTMRs terminate in Merkel discs of touch domes, respond preferentially to skin indentation, and report on the static nature of tactile stimulations (Koltzenburg et al., 1997; Woodbury and Koerber, 2007). Although the morphology, physiology, and function of F-LTMRs are less well understood and they are not yet genetically identified, in cats they display A β conduction velocities, exhibit large receptive fields, and while they are highly sensitive to stroking of hairy skin, they respond poorly to skin indentation and deflection of individual hairs. A fourth hairy skin LTMR type, A δ -LTMRs, are the most sensitive of the LTMRs, have lightly myelinated axons with an intermediate conduction velocity and, like A β RA-LTMRs, they are velocity detectors that rapidly adapt to sustained stimulation. Originally thought to associate exclusively with small hair types, termed down hairs in cats, it is now established that A δ -LTMR responses are elicited following movement of multiple hair types (Brown and Iggo, 1967; Burgess et al., 1968; Horch et al., 1977). Finally, C-LTMRs exhibit a slow conduction velocity and an intermediate rate of adaptation, and recent work in humans suggests an involvement in pleasurable or emotional touch because they are optimally tuned to stroking of the skin at rates that are deemed pleasurable (Hamann, 1995; Horch et al., 1977; Olausson et al., 2002). While each of the main hairy skin LTMRs is sensitive to innocuous touch of the skin or body hairs, the mechanisms by which LTMR subtypes are differentially tuned to specific-touch stimuli remains incompletely understood. Thus, we previously employed a molecular genetic labeling strategy to study peripheral and central axonal projections of distinct LTMR subtypes with the goal of uncovering morphological correlates of LTMR subtype response properties and the functional organization of LTMR projections (Li et al., 2011). The peripheral terminals of A β RA-LTMRs, A δ -LTMRs, and C-LTMRs in murine back hairy skin form lanceolate axonal endings in unique combinations with the three main types of hair follicles of the mouse; guard, awl/auchene, and zigzag hairs. Guard hair follicles, the largest and least abundant, comprising ~1% of all murine hairs, receive rich innervation by A β RA-LTMR lanceolate endings and also associate with A β SAI-LTMR endings, which terminate upon Merkel discs in touch dome complexes. Awl/auchene hairs make up ~20% of hair follicles and are triply innervated by interdigitated A β RA-LTMR, A δ -LTMR, and C-LTMR lanceolate endings. The most abundant and smallest hairs, zigzag hairs, comprise ~80% of hair follicles and are innervated by interdigitated A δ -LTMR and C-LTMR lanceolate endings. Each of the three hair follicle types is also surrounded by circumferential endings belonging to neurons of unknown physiological properties. Thus, each of the three distinct hair types is associated with unique combinations of LTMR axonal endings, thereby rendering them neurophysiologically distinct (Li et al., 2011).

The contributions of LTMR subtypes and downstream spinal cord and brainstem circuit components to the capture, processing, and perception of the direction of stimulus movement across the skin are not known. LTMRs that innervate hairy skin are themselves candidates for having direction-selective tuning properties because each hairy skin LTMR subtype is associated with one or more hair follicle types and several are highly sensitive to hairy skin stroking and hair deflection. Moreover, cat, rat,

and mouse trigeminal mechanosensory neurons that associate with whisker follicles of mystacial pads exhibit direction-selective responses to whisker deflection (Gottschaldt and Vahle-Hinz, 1981; Kwegyir-Afful et al., 2008; Lichtenstein et al., 1990). On the other hand, attempts to address the issue of direction selectivity of hairy skin LTMRs of cats, rabbits, and primates have suggested that LTMRs are (Brown and Iggo, 1967; Maruhashi et al., 1952; Tuckett, 1978) or are not (Essick and Whitsel, 1985; Greenspan, 1992; Hyvärinen and Poranen, 1978; Whitsel et al., 1972) sensitive to the direction of deflection of normal hair follicles. One study that asked whether feline hairy skin A δ -LTMRs exhibit direction-selective responses concluded that these neurons may be direction-selective but that it is difficult to ascertain because of their ultrasensitive response property (Ray et al., 1985). Indeed, no one LTMR subtype has been unequivocally shown to be differentially sensitive to the direction of movement of objects across hairy skin, leading to the idea that information about direction of stimulus movement is represented exclusively by the firing patterns evoked in populations of mechanoreceptors activated by a moving tactile stimulus, rather than by direction selectivity of individual LTMR subtypes (Essick and Edin, 1995). Thus, whether A δ -LTMRs, the most sensitive of hairy skin mechanoreceptors, or other hairy skin LTMR subtypes exhibit direction-selective responses to hair deflection or skin stroking remains an open question, and quantitative measures of LTMR response properties in conjunction with sensitive hair deflection paradigms are needed to address this question.

Here, using novel methods for deflecting individual hair follicles in the four cardinal directions relative to the body axis of the mouse, we report that A δ -LTMRs are more sensitive to hair deflection in the caudal-to-rostral (R) direction than in the rostral-to-caudal (C) direction. A morphological correlate of this unique tuning property is a striking enrichment of A δ -LTMR lanceolate endings on the caudal side of awl/auchene and zigzag hair follicles. Interestingly, expression of the neurotrophic growth factor BDNF, which is well known to control axonal growth and target innervation, is localized to caudal side epithelial cells of these hair follicles, the BDNF receptor TrkB is highly expressed in A δ -LTMRs, and conditional ablation of *BDNF* in skin epithelial cells leads to a loss of A δ -LTMR lanceolate ending polarization. As a result of loss of lanceolate ending polarization, direction-selective tuning of A δ -LTMRs to hair deflection is randomized in these conditional BDNF mutant mice. Thus, A δ -LTMRs are direction-selective hairy skin mechanoreceptors, and BDNF-to-TrkB signaling between hair follicle epithelial cells and A δ -LTMRs lanceolate endings establishes A δ -LTMR terminal polarization and their direction selectivity to hair deflection.

RESULTS

A δ -LTMRs Are Tuned to the Direction of Hair Deflection

We previously established molecular-genetic tools to investigate A β RA-, A δ -, and C-LTMRs and demonstrated that these sensory neuron subtypes form interdigitated longitudinal lanceolate endings in close association with select types of hair follicles in mouse hairy skin. Utilizing a *TrkB^{GFP}* knockin mouse line, TrkB⁺ DRG neurons of adult mice were found to be A δ -LTMRs, which comprise ~7% of adult thoracic DRG neurons and exhibit

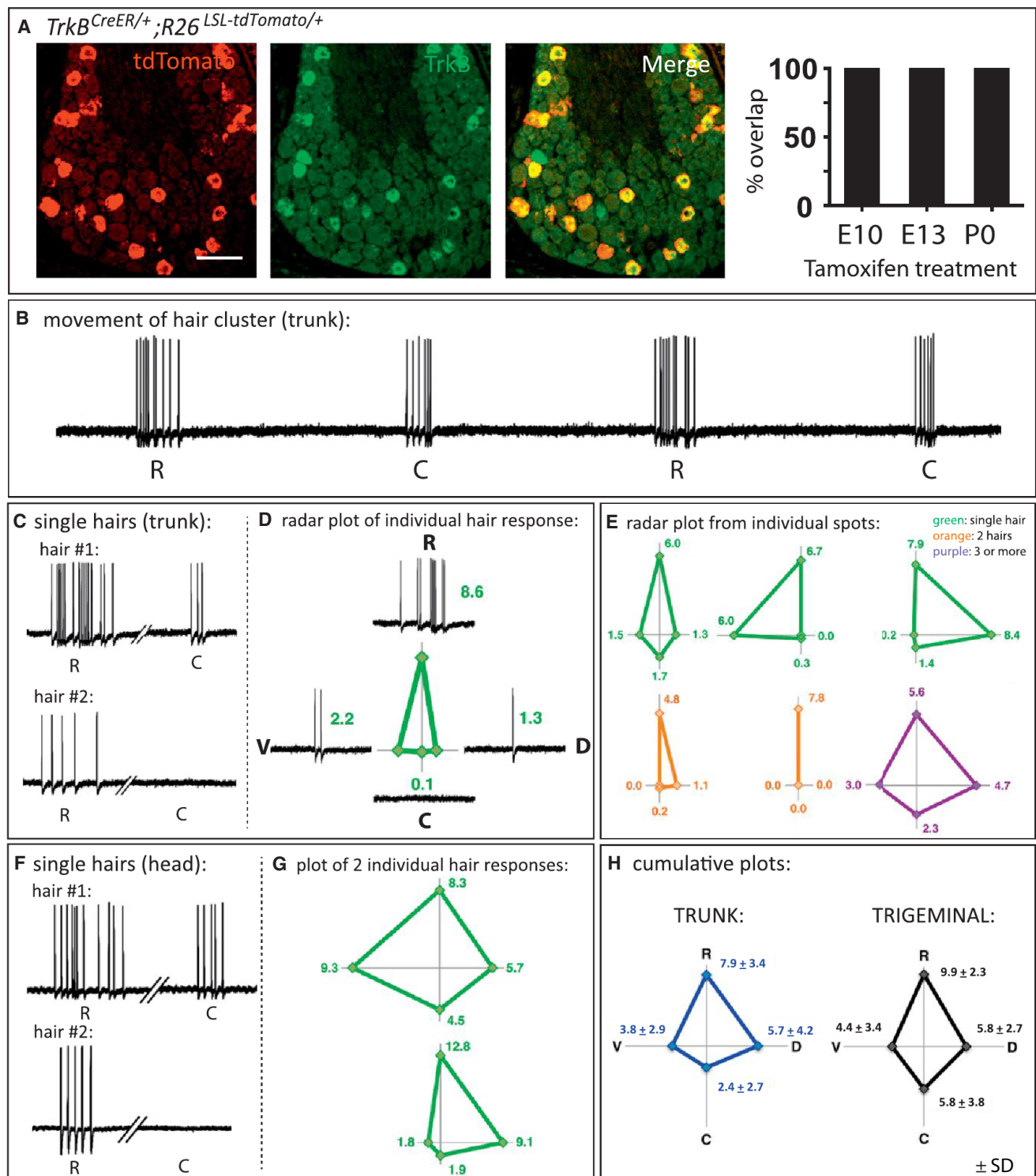


Figure 1. A δ -LTMRs Exhibit Direction-Selective Tuning to Hair Follicle Deflection

(A) Pregnant *TrkB^{CreER/+};Rosa26^{LSL-tdTomato/+}* mice were treated with tamoxifen (3 mg at either E10, E13, by oral gavage or a single intraperitoneal [i.p.] injection of 1mg at P0). Double fluorescence in situ hybridization for tdTomato (red) and TrkB (green) on thoracic DRG sections from P21 mice reveals that virtually all tdTomato+ DRG neurons are TrkB+ (100%, 194/194). A small subset of TrkB+ cells observed by in situ hybridization (green) are not labeled with the *TrkB^{CreER}* labeling strategy (3/27 in this particular image). Scale bar represents 100 μ m.

(B) Intracellular ex vivo electrophysiological recordings from a *TrkB^{GFP}*-labeled A δ -LTMR reveal selectivity of response to repetitive stimulation of a small cluster of hairs in the rostrocaudal direction; note the consistently greater response elicited by deflections in the rostral (R) versus caudal (C) direction.

(legend continued on next page)

central axonal projections that terminate within lamina II/III of the spinal cord dorsal horn. The peripheral projections of A δ -LTMRs form longitudinal lanceolate endings associated with both awl/auchene and zigzag hairs of trunk hairy skin (Li et al., 2011). To facilitate examination of the physiological and morphological properties of individually labeled A δ -LTMRs, we generated a *TrkB^{CreER}* knockin mouse line in which a fusion cassette consisting of Cre recombinase and a triple mutant form of the human estrogen receptor (CreERT2) was introduced via homologous recombination into the first coding exon of the *TrkB* gene (Figure S1 available online). Treatment of *TrkB^{CreER}*; *Rosa26^{LSL-tdTomato}* or *TrkB^{CreER}*; *Rosa26^{iAP}* mice (Badea et al., 2009; Madisen et al., 2010) with a single injection of 4-hydroxytamoxifen (4-HT) at embryonic day 10 (E10) or E13 led to constitutive expression of tdTomato or alkaline phosphatase (AP), respectively, in medium diameter DRG neurons of adult mice, which express TrkB ($n = 194$ cells labeled in *TrkB^{CreER}*; *Rosa26^{LSL-tdTomato}* mice; Figure 1A). As expected, the peripheral axons of sparsely labeled neurons form longitudinal lanceolate endings at awl/auchene and zigzag hairs. Their central projections terminate in bouton-rich “flame-shaped arbors” that are continuous throughout the rostral-caudal axis of laminae II/III of the spinal cord dorsal horn (Figure S2E). The longitudinal extent of A δ -LTMR central terminations range from 250 μ m in cervical and lumbar regions to 500 μ m in thoracic segments (Figures S2A–S2D). In addition, sharp electrode recordings of *TrkB^{CreER}*-labeled neurons in intact ex vivo and in vivo preparations revealed that they exhibit A δ -LTMR physiological properties, with narrow uninflected somal spikes, A δ conduction velocities, low mechanical thresholds, and rapidly adapting responses to indentation of the skin (Figures S2G and S2H). Thus, *TrkB^{CreER}* mice specifically label A δ -LTMRs enabling versatile genetic access to this neuronal population.

Because A δ -LTMRs terminate in lanceolate endings associated with hair follicles, we next sought to define their response properties with respect to hair follicle deflection. As shown previously in cats, rabbits, rats, and mice, A δ -LTMRs are exquisitely sensitive and exhibit robust, rapidly adapting responses to maintained stimuli. These neurons respond throughout the dynamic phase of stimuli, providing bursts of spikes at both the onset and termination of sustained mechanical stimuli. To investigate A δ -LTMR responses to hair deflection, we used fine probes to deflect small clusters of 3–6 hairs within the receptive fields of *TrkB^{GFP}*-labeled A δ -LTMRs. As observed with sustained indentation, A δ -LTMRs responded briskly throughout the movement of hairs but adapted rapidly when movement ceased (Figure 1B). Remarkably, cells were found to respond more strongly to deflection of groups of hairs in the caudal-to-

rostral (R) direction, compared to the rostral-to-caudal (C) direction.

Because deflection of hair clusters subjected individual hairs within the clusters to unpredictable movements, we next asked whether direction-selective responses of A δ -LTMRs are observed following controlled deflection of individual hairs. High-resolution mapping was performed throughout the cell's receptive field, using controlled deflections in the four cardinal directions, as clusters were successively refined down to a single hair. These fine-grained analyses, akin to fiber-teasing techniques in extracellular recordings from nerve bundles, revealed that the large receptive fields of A δ -LTMRs represent a mosaic patchwork; movements of doublets and triplets elicited responses that were often indistinguishable from single hairs; conversely, large numbers of individual hairs within the boundaries of receptive fields often elicited no response in the cell (data not shown) despite being adjacent to and/or surrounded by hairs that did, suggesting remarkably little sensory coupling between adjacent hair follicles in the dermis.

As observed for A δ -LTMRs responses to deflection of groups of hairs, a comparison of A δ -LTMR responses to deflection of individual hairs revealed a pronounced direction-selective tuning property. Optimal A δ -LTMR responses were generally elicited by movement in the R direction; for the majority of single hairs and doublets tested, R tuning was extremely sharp, and responses to movement in the C direction were frequently nonexistent. Deflections in the orthogonal plane revealed that many hairs also elicited good responses following deflection in the ventral-to-dorsal (D) direction or alternatively, the dorsal-to-ventral (V) direction, in addition to the R direction, suggesting that our manipulator may not have been aligned with the optimal vector for that hair (Figure 1E). Combining the responses from all spots tested throughout receptive fields of A δ -LTMRs, from single hairs to groups of five, revealed strong directional tuning overall; the average number of spikes elicited in A δ -LTMRs by R deflections of hairs was over 3-fold more than the number elicited from C deflections of the same magnitude ($p < 0.01$, Figure 1H). Finally, we performed similar analyses using an in vivo preparation that enabled recordings of trigeminal ganglion *TrkB^{CreER}*-labeled A δ -LTMRs, which innervate hairy skin of the head. Although facial hairs exhibited greater heterogeneity in orientation than trunk hairs, trigeminal A δ -LTMRs also exhibit direction-selective patterns of activation following hair deflection (Figures 1F and 1G). Direction selectivity in response to hair deflection is not a property of all LTMR subtypes because in vivo recordings of trigeminal A β RA-LTMRs, which form lanceolate endings associated with guard and awl/auchene hair follicles,

(C) Examples of responses to deflection of three different single hairs in the receptive fields of A δ -LTMRs in *K5^{Cre/+}*; *BDNF^{flax/+}*; *TrkB^{GFP/+}* mice.

(D) Vector plot of average responses to multiple deflections of a single hair in the four cardinal directions.

(E) Examples of different vector plots elicited by movements of hairs in the receptive fields of A δ -LTMRs in *K5^{Cre/+}*; *BDNF^{flax/+}*; *TrkB^{GFP/+}* mice (green, single hairs; orange, hair doublets; purple, cluster).

(F) In vivo electrophysiological recordings of A δ -LTMRs in trigeminal ganglia of *TrkB^{CreER/+}*; *Rosa26^{LSL-tdTomato/+}* mice showing two different examples of responses elicited by controlled deflections of single hairs.

(G) Examples of vector plots of average responses to multiple deflections of single hairs in the four cardinal directions in trigeminal A δ -LTMRs in vivo.

(H) Cumulative vector plot of all individual locations analyzed throughout the receptive fields of trunk and trigeminal A δ -LTMRs in vivo, revealing a similar tuning for rostral deflections of hairs as in trunk DRG. Shown are the means \pm SD.

See also Figures S1, S2, and S3.

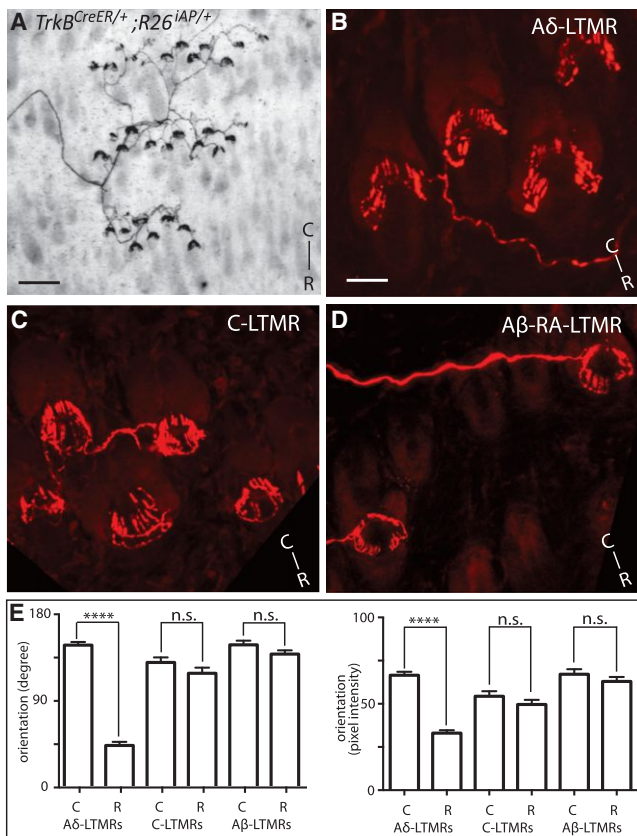


Figure 2. Aδ-LTMRs Endings Are Polarized around Hair Follicles

(A) Visualization of Aδ-LTMR peripheral receptive fields. *TrkB^{CreER/+}; Rosa26^{iAP/+}* mice were treated with tamoxifen (1 mg at E13 or P25, by oral gavage) and whole mount, alkaline phosphatase preparation performed on skin of P21 or P40–P50 animals. Each Aδ-LTMR was found to arborize and form longitudinal lanceolate endings associated an average of 35 zigzag and awl/auchene hair follicles. Scale bar represents 100 μm.

(B) Aδ-LTMR longitudinal lanceolate endings are polarized toward the caudal side of hair follicles. Whole mount immunohistochemical staining showing Aδ-LTMR longitudinal lanceolate endings more densely populated on the caudal (top of the image) side (note: *TrkB^{CreER/+}; Rosa26^{LSL-tdTomato/+}* mice treated with 1 mg tamoxifen at E13; n = 4). Scale bar represents 20 μm.

(C and D) There is no obvious polarization of C-LTMR and Aβ RA-LTMR longitudinal lanceolate endings. Whole mount immunohistochemical staining of sparsely labeled C-LTMR lanceolate endings in *TrkB^{CreER/+}; Rosa26^{LSL-tdTomato/+}* mice treated with 2 mg tamoxifen/day at P13–P15, n = 3. Whole mount immunohistochemical staining of sparsely labeled Aβ RA-LTMR lanceolate endings in *Ret^{CreER/+}; Rosa26^{LSL-tdTomato/+}* mice treated with 1 mg tamoxifen/day at E10 and E11 (n = 3).

(E) Quantification of orientation/morphological properties of Aδ-, C-, and Aβ RA-LTMR longitudinal lanceolate endings. Shown are means ± SEM. Left: illustration of central angle measurement to compare the density of lanceolate endings on caudal versus rostral side of the hair follicle; Right: summary of pixel intensity quantification by the ImageJ software. Lanceolate ending polarity was assessed by central angle and pixel intensity measurement on mice with sparsely labeled Aδ-, C-, and Aβ RA-LTMRs, respectively (n = 3). ****p < 0.001.

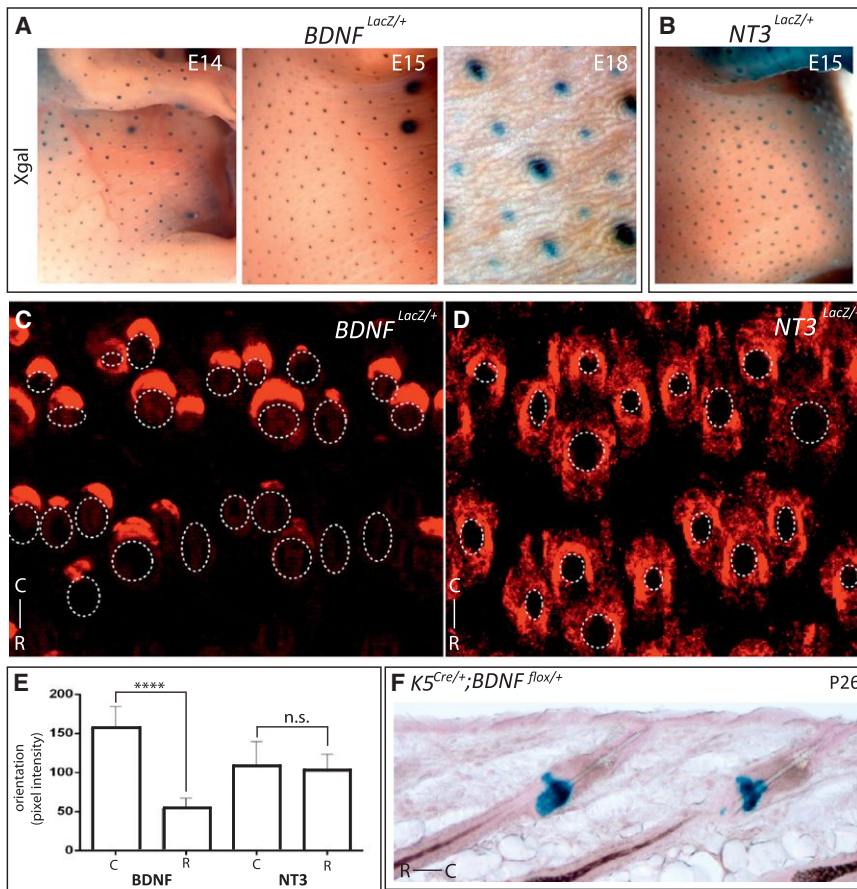
failed to show a directional preference (Figure S3). Thus, DRG and trigeminal ganglion Aδ-LTMRs are ultrasensitive mechanosensory neurons optimally tuned to hair movement in the R direction (Figure 1H).

Aδ-LTMRs Endings at Hair Follicles Are Polarized

In the visual system, a small subset of direction-selective retinal ganglion cells (DS-RGCs) exhibit polarized dendrites that are oriented in the direction of their optimal responses, and this morphological feature may underlie direction-selective capture of visual images moving across the receptive fields of this subset (Kim et al., 2008; Trenholm et al., 2011). To begin to ask whether there is a morphological basis of Aδ-LTMR direction-selective tuning to hair deflection, the peripheral axonal ending morphology of individual Aδ-LTMRs was visualized. To accomplish this, we used *TrkB^{CreER}; Rosa26^{iAP}* and *TrkB^{CreER}; Rosa26^{LSL-tdTomato}* mice and induced expression of AP or tdTomato, respectively, in small numbers of Aδ-LTMRs thereby enabling detailed examination of their cutaneous axonal morphologies. Sections of back hairy skin collected from *TrkB^{CreER}; Rosa26^{iAP}* and *TrkB^{CreER}; Rosa26^{LSL-tdTomato}* mice were analyzed using either whole mount AP or immunohistochemical staining techniques. The peripheral axonal branches of individual adult Aδ-LTMRs were found to arborize and form longitudinal lanceolate endings associated an average of 35 zigzag and awl/auchene hair follicles, encompassing a skin area of 0.6–0.8 mm² (n = 3; Figure 2A). Interestingly, both low and high magnification images of Aδ-LTMR projections in the skin revealed a marked polarization of their lanceolate endings surrounding hair follicles. These lanceolate endings are greatly enriched on the caudal side of hair follicles (Figure 2B). Lanceolate ending polarization is unique to Aδ-LTMRs as neither C-LTMRs nor Aβ RA-LTMRs exhibit appreciable polarization (Figures 2C and 2D). Vector analysis reveals a highly significant polarization of Aδ-LTMR endings on the caudal sides of awl/auchene and zigzag follicles in back hairy skin (Figure 2E). This polarization may provide a structural basis of selective responses of Aδ-LTMRs to hair deflection in the R direction.

BDNF Expression Is Polarized and Concentrated on the Caudal Side of Hair Follicles in Close Association with Aδ-LTMR Lanceolate Endings

We next sought to identify asymmetrically localized cue(s) that instruct the polarization of Aδ-LTMR lanceolate processes around hair follicles because ablation of such a cue, as a means to disrupt lanceolate ending polarization, would allow us to ask whether lanceolate ending polarization underlies Aδ-LTMR direction-selective responses. Our analysis focused on the temporal and spatial patterns of expression of the neurotrophins in hairy skin during Aδ-LTMR development because robust expression of the neurotrophin receptor TrkB is a distinguishing feature of Aδ-LTMRs and because neurotrophins are well known to control sensory axon development and target field innervation. TrkB has two main ligands, brain-derived neurotrophic factor (BDNF) and neurotrophin-4 (NT4), while a third neurotrophin, neurotrophin-3 (NT3), can bind TrkB with low affinity (Klein et al., 1991, 1992). To examine the expression patterns of BDNF, NT4, and NT3 in the context of hair follicle development and sensory neuron innervation, we used *BDNF^{LacZ}*, *NT3^{LacZ}*, and *NT4^{LacZ}* knockin mouse lines (Fariñas et al., 1994; Gorski et al., 2003; Liu et al., 2012). Whole-mount X-gal staining of *BDNF^{LacZ}* and *NT3^{LacZ}* embryos revealed robust LacZ expression associated with developing hair follicles in



animals of both genotypes, whereas $NT4^{LacZ}$ mice failed to exhibit appreciable expression. LacZ expression in both $BDNF^{LacZ}$ and $NT3^{LacZ}$ embryos, but not $NT4^{LacZ}$ embryos, was detected during the first wave of hair follicle morphogenesis (~E14) and persisted through the second (E15) and third waves (E18) (Figures 3A, 3B, S4A, and S4B). Hair follicle-associated BDNF and NT3 expression was also observed in adulthood (data not shown). Strikingly, asymmetrically localized BDNF concentrated on the caudal side of hair follicles was observed just below the sebaceous gland in a bulbous region of the follicle, in close proximity to A δ -LTMR lanceolate endings (Figure 3C). To determine whether BDNF is produced in hair follicle cells of epithelial origin, mice in which Cre recombinase is expressed in skin keratinocytes ($K5^{Cre}$) were crossed with mice harboring a $BDNF^{flox}$ conditional reporter allele (Gorski et al., 2003; Ramirez et al., 2004). Thus, following Cre-mediated excision, one copy of BDNF is excised and expression of a functional LacZ reporter cassette becomes activated. Indeed, robust, asymmetrically localized LacZ staining was observed in $K5^{Cre}; BDNF^{flox}$ mice (Figures 3F and S4C), indicating that BDNF is produced in epithelial cells of the caudal region of hair follicles. In contrast, NT3 exhibits a symmetric pattern of expression around zigzag and awl/auchene hair follicles, although it is asymmetrically expressed around guard hairs, which are not innervated by A δ -LTMRs (Figures 3D and 3E). Thus, while both BDNF and NT3 are expressed in epithelial cells

Figure 3. BDNF Expression Is Highly Polarized to the Caudal Side of Hair Follicles

(A and B) Embryonic $BDNF^{LacZ/+}$ and $NT3^{LacZ/+}$ tissue was examined by whole mount Xgal staining. Robust BDNF-LacZ and NT3-LacZ expression was detected in developing hair follicles throughout all three stages of hair follicle morphogenesis. (C and D) Immunostaining of β gal (red) on back hairy skin sections from $BDNF^{LacZ/+}$ and $NT3^{LacZ/+}$ P7 mice shows that BDNF-LacZ expression is polarized while NT3-LacZ expression is not. (E) Summary of pixel intensity quantification by the ImageJ software. LacZ expression polarity was assessed by pixel intensity measurement on P7 mice. Shown are means \pm SEM. **** $p < 0.001$. (F) Xgal staining on $K5^{Cre/+}; BDNF^{flox/+}$ embryos shows that hair follicle-derived BDNF originates in epithelial cells. See also Figure S4.

of awl/auchene and zigzag hair follicles, BDNF, but not NT3, is asymmetrically distributed in epithelial cells surrounding these follicles being greatly enriched on the hair follicle caudal side, which harbors A δ -LTMR endings.

To directly compare the spatial and temporal relationships between BDNF expression and nascent A δ -LTMR lanceolate endings associated with hair follicles during development, we next generated $BDNF^{LacZ}; TrkB^{CreER}; Rosa26^{LSL-tdTomato}$ mice. These mice were given a low-dose of tamoxifen (1 mg) at E13.5 to enable simultaneous visualization of BDNF expression patterns and nascent A δ -LTMR axonal endings around hair follicles. A δ -LTMR axons reach the skin prior to birth and associate with hair follicles at neonatal times (data not shown). Subsequently, beginning approximately postnatal day 1 (P1), lanceolate processes emerge from rudimentary processes, circumferentially oriented around hair follicles. By P7, fully formed A δ -LTMR lanceolate endings concentrated on the caudal sides of hair follicles are observed. Interestingly, during the time of A δ -LTMR axonal ending growth and maturation, the locations of lanceolate processes and BDNF expression is strikingly coincident. At P5, when nascent lanceolate endings are extending along the longitudinal axis of hair follicles, they are intimately associated with hair follicle epithelial cells that express BDNF and not with epithelial cells that lack BDNF (Figure 4). Thus, hair follicle epithelial cell-derived BDNF is temporally and spatially positioned to serve as a growth and guidance cue for A δ -LTMR lanceolate endings during their period of extension and polarization around hair follicles.

BDNF Expression in Hair Follicle Epithelial Cells Is Required for Polarization of A δ -LTMR Endings

The robust expression of BDNF in caudally located hair follicle epithelial cells that are in close proximity to developing

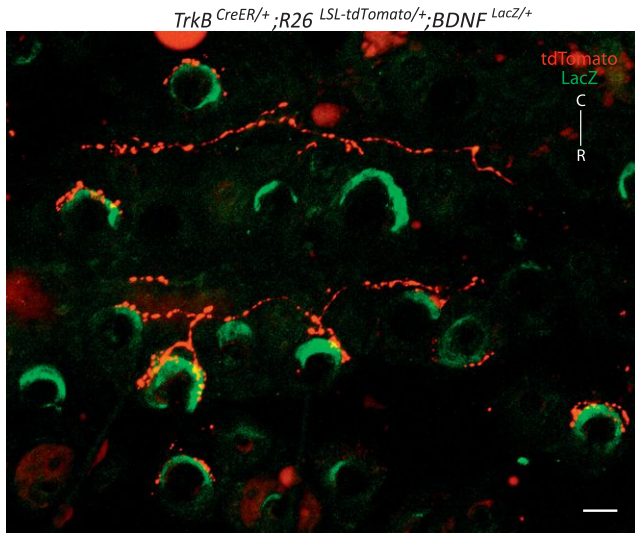


Figure 4. BDNF Expression Is Temporally Coincident with and Occurs in Close Apposition to Developing A δ -LTMRs Longitudinal Lanceolate Endings

In hairy skin sections from P5 *TrkB^{CreERT2/+};Rosa26^{LSL-tdTomato/+};BDNF^{LacZ/+}* mice, A δ -LTMR endings axonal endings are labeled with tdTomato fluorescence (red) and BDNF is labeled by β gal immunostaining (green). Cre recombinase was induced by administration of 1 mg tamoxifen at E13 (oral gavage, $n = 2$). Note that postnatal, BDNF-producing epithelial cells are in close apposition to A δ -LTMR endings. Scale bar represents 30 μ m.

lanceolate endings, and of TrkB in A δ -LTMRs, suggests a role for BDNF-TrkB signaling in A δ -LTMR lanceolate ending polarization. Therefore, we next asked whether TrkB and its ligand BDNF control development of A δ -LTMRs and the extent to which BDNF-TrkB signaling contributes to A δ -LTMR lanceolate ending polarity. Indeed, through monitoring GFP expression from the *TrkB^{GFP}* allele as a general readout of the integrity of A δ -LTMRs, a nearly complete loss of GFP⁺ neurons in *TrkB^{GFP}* homozygous null mutants at the day of birth (P0) was observed. Likewise, BDNF null mutants exhibited a 66% reduction in the number of GFP⁺ A δ -LTMRs (Figures S5A and S5B). Thus, embryonic BDNF-TrkB signaling is necessary for A δ -LTMR development. We next asked whether BDNF produced in hair follicle epithelial cells contributes to A δ -LTMR development. The *TrkB^{GFP}* allele was again used to monitor A δ -LTMR integrity, and the number of thoracic level GFP⁺ DRG neurons in *BDNF^{f/f}* mice expressing Cre recombinase in each of three general cell types that normally express this ligand were counted. To conditionally ablate BDNF in epithelial cells, *K5^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}* were generated. The number of GFP⁺ neurons in adult *K5^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}* were comparable to their littermate controls (Figures S5C and S5D). In contrast, mesenchyme-derived BDNF is required for A δ -LTMR development because a 66% reduction of GFP⁺ neurons was observed in P0 *T^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}* animals, which express Cre recombinase in cells derived from mesoderm (Perantoni et al., 2005) (Figure S5D). *Wnt1^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}*, which lack BDNF in all DRG neurons and Schwann cells, exhibited a normal complement of A δ -LTMRs. These findings indicate that BDNF

emanating from the mesenchyme, and not hair follicle epithelial cells or DRG neurons themselves or their associated Schwann cells, is essential for maturation and general development of A δ -LTMRs.

The normal complement of A δ -LTMRs found in *K5^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}* mice indicates that epithelial cell-derived BDNF is dispensable for general maturation and survival of these neurons. Therefore, we next used *K5^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}* mice to assess the role of BDNF expressed in hair follicle-associated epithelial cells for hair follicle innervation and polarization of A δ -LTMR lanceolate endings. While polarization of A δ -LTMR lanceolate endings was readily apparent in control animals, lanceolate endings of *K5^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}* mice, while present, showed a marked loss of polarization (Figures 5A and 5B). Some hair follicles exhibited a bias of endings on one side or the other, but the overall pattern of lanceolate ending organization was randomized with respect to hair follicle orientation (Figure 5C). Thus, while hair follicle epithelial cell-derived BDNF is dispensable for general maturation and survival of A δ -LTMRs, and for hairy skin innervation, it is essential for morphological polarization of A δ -LTMR lanceolate endings on the caudal sides of awl/auchene and zigzag hair follicles.

BDNF Expression in Hair Follicle Epithelial Cells Is Required for Direction-Selective Responses of A δ -LTMRs

The lack of A δ -LTMR lanceolate ending polarization in *K5^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}* mice renders these animals valuable for asking whether morphological polarization of lanceolate endings underlies direction-selective tuning of A δ -LTMRs to hair deflection. Therefore, we next used *K5^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}* mice for ex vivo skin-nerve recordings to address this possibility. Remarkably, in *K5^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}* mice, A δ -LTMR responses to oriented hair deflections were randomized with respect to the direction of hair deflection (Figure 6B). While many individual hairs in the receptive fields of A δ -LTMRs in *K5^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}* mice exhibited preferential tuning to deflection in the R direction, as in neurons from wild-type mice, others showed preference to deflection in the opposite, C direction; still others showed preferential tuning to directions of hair movement that were not seen among hairs innervated by A δ -LTMRs in wild-type mice (Figures 6C and 6D). Quantification of the results from fine-grained analyses throughout the receptive fields of A δ -LTMRs showed that responses in the mutants were randomized with respect to direction of hair deflection (Figure 6E). On the other hand, the somal spikes, peripheral conduction velocities, adaptation properties to sustained stimuli, sensitivity to cooling, and mechanical thresholds of A δ -LTMRs to stimulation with von Frey filaments were normal in *K5^{Cre}; BDNF^{flox/flox}; TrkB^{GFP}* (Figure S6). Thus, while A δ -LTMR sensitivity and basic physiological properties are intact in the absence of hair follicle epithelial cell-derived BDNF, they fail to exhibit direction-selective responses to hair follicle deflection. We conclude that the morphological polarization A δ -LTMR lanceolate endings associated with hair follicles underlies direction-selective tuning of A δ -LTMRs to hair deflection.

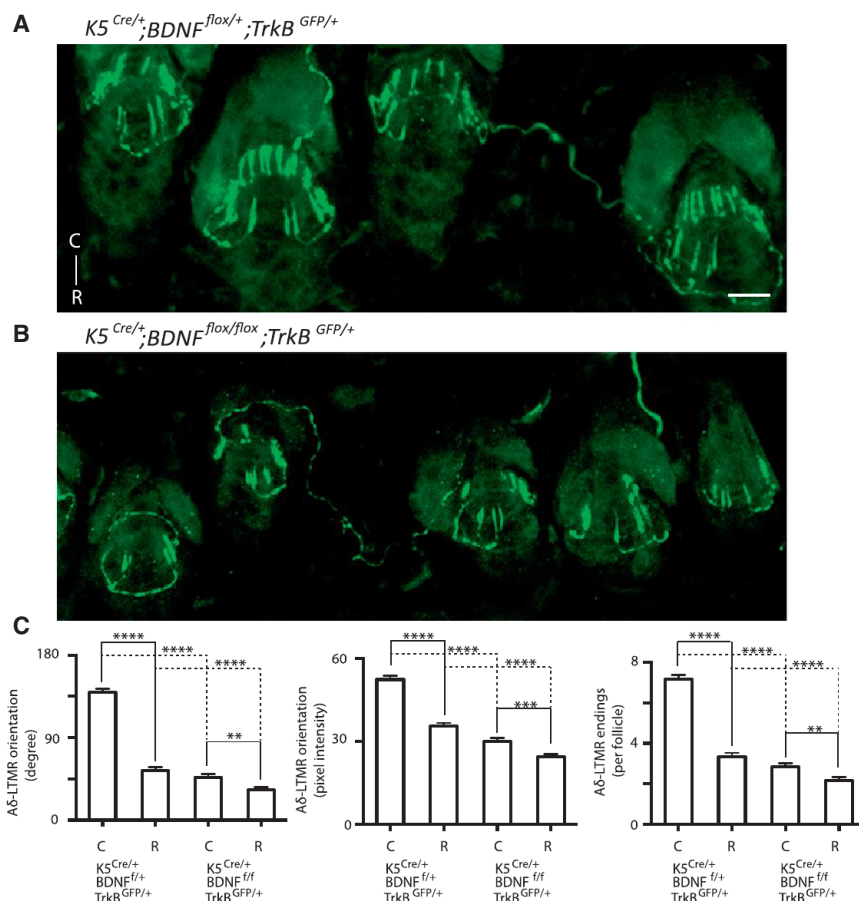


Figure 5. BDNF Expression in Hair Follicle Epithelial Cells Is Required for Polarization of Aδ-LTMR Endings

(A and B) Whole mount skin staining in $K5^{Cre/+}; BDNF^{flox/+}; TrkB^{GFP/+}$ and $K5^{Cre/+}; BDNF^{flox/flox}; TrkB^{GFP/+}$ mice. Aδ-LTMR lanceolate endings (GFP, green) are polarized on the caudal side of hair follicles in littermate control $K5^{Cre/+}; BDNF^{flox/+}; TrkB^{GFP/+}$ mice (A). Loss of Aδ-LTMR lanceolate ending polarity in $K5^{Cre/+}; BDNF^{flox/flox}; TrkB^{GFP/+}$ mice (B) (endings associated with 106 hair follicles; $n = 2$ mice).

(C) Aδ-LTMR lanceolate ending polarity is dramatically reduced in $K5^{Cre/+}; BDNF^{flox/flox}$ mice. Lanceolate ending polarity was assessed by central angle and pixel intensity measurement and quantification of lanceolate endings in $K5^{Cre/+}; BDNF^{flox/flox}; TrkB^{GFP/+}$ and littermate control $K5^{Cre/+}; BDNF^{flox/+}; TrkB^{GFP/+}$ mice. 106 hair follicles were quantified for each genotype ($n = 2$). Scale bar represents 20 μ m. See also Figure S5.

cells surrounding the hair follicle. While the polarized expression of BDNF in hair follicle epithelial cells is crucial for Aδ-LTMR ending polarity, and thus the direction-selective tuning property that is unique to this LTMR subtype, it is dispensable for maturation and survival of these neurons. In contrast, mesenchyme-derived BDNF is essential for the general development of Aδ-LTMRs. Thus, distinct sources of BDNF contribute

DISCUSSION

Direction-selective responses to sensory stimuli is a hallmark feature of several sensory systems, and elucidating the mechanisms underlying direction selectivity, from neurons to circuits, is essential for understanding how we perceive our environment. Here, we report that a subtype of primary somatosensory neurons, Aδ-LTMRs, are tuned to the direction of hair deflection as a result of developmental mechanisms dictating morphological features unique to this LTMR subtype. Aδ-LTMRs respond more strongly to hair deflection in the R direction than in the C direction. This tuning property results from polarization of Aδ-LTMR lanceolate endings, which are concentrated on the caudal side of hair follicles. BDNF from epithelial cells signaling through TrkB expressed in Aδ-LTMRs directs development of their lanceolate ending morphological polarization. Elimination of BDNF in hair follicle epithelial cells leads to loss of both morphological polarization and direction-selective tuning. Thus, polarized expression of BDNF in epithelial cells on the caudal sides of hair follicles underlies direction-selective tuning of Aδ-LTMRs.

Polarized expression in caudally located hair follicle epithelial cells is a distinguishing feature of BDNF because the related neurotrophic factor NT3, which is also transcribed in a subset of hair follicle epithelial cells, is expressed in a pattern that is not polarized. NT3 is expressed in a circumferential manner, in epithelial

to different aspects of Aδ-LTMR development and function. Key to understanding the establishment of Aδ-LTMR lanceolate ending polarization, and thus direction selectivity of Aδ-LTMRs, is elucidating the mechanism of polarized BDNF transcription in hair follicle epithelial cells. It is possible that the same cues that govern polarization of other hair follicle features, such as the location of sebaceous glands and a variety of molecular markers, also control polarized expression of BDNF and thus Aδ-LTMR lanceolate ending polarization and their direction selectivity to hair deflection.

Of the many questions pertaining to direction-selective Aδ-LTMRs, a most intriguing one is how these ultrasensitive mechanosensory neurons respond preferentially to deflection of individual hairs in the R direction compared to the C direction. Aδ-LTMR longitudinal lanceolate endings associate with the caudal sides of hair follicles. Each neuron's cutaneous projection branches within the skin to innervate on average 35 awl/auchene and zigzag hair follicles, with 10–20 finger-like lanceolate endings closely associated with each follicle. Surrounding each lanceolate ending are processes of terminal Schwann cells (TSCs), which together with the lanceolate ending and hair follicle epithelial cells constitute "lanceolate complexes." Ultrastructurally, each lanceolate complex exhibits gaps or openings in which the Aδ-LTMR axonal membrane on the side facing the hair follicle epithelial cell is exposed, forming a

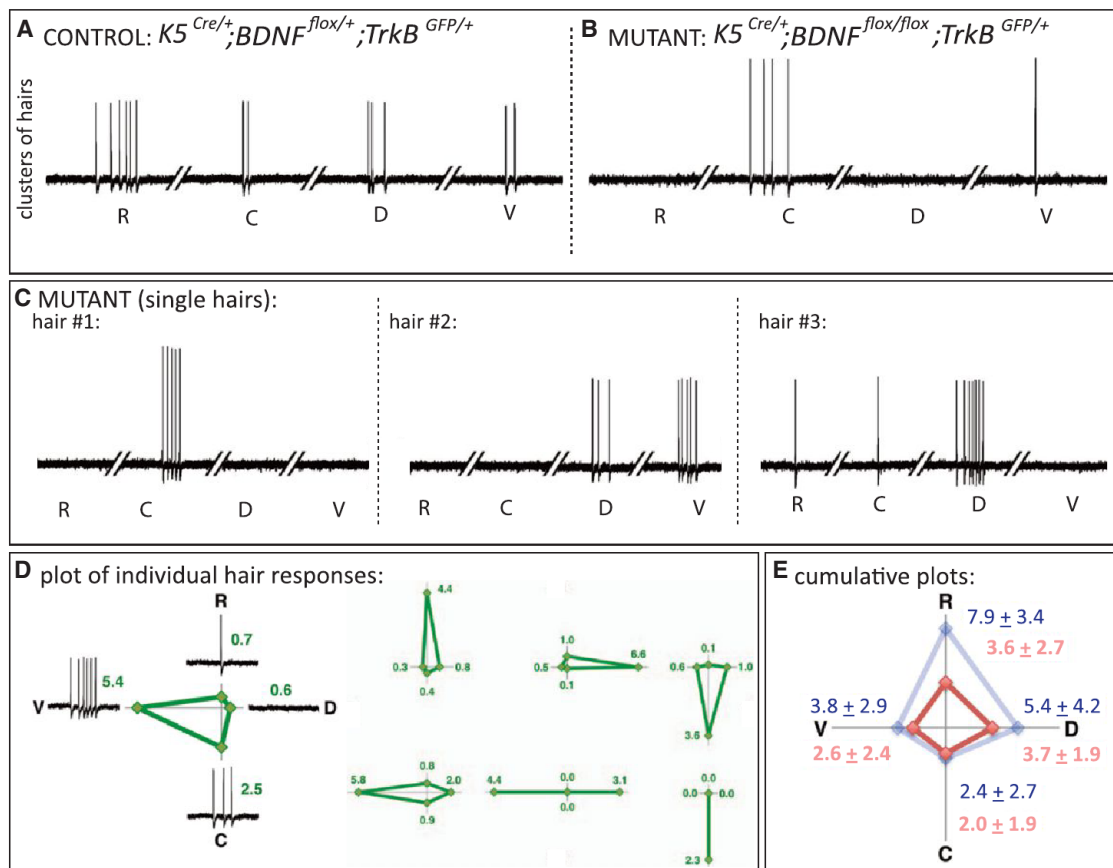


Figure 6. BDNF Expression in Hair Follicle Epithelial Cells Is Required for Direction-Selective Tuning of A δ -LTMRs

(A and B) Ex vivo intracellular recordings of A δ -LTMR responses elicited by deflecting a small cluster of hairs in (A) $K5^{Cre/+}; BDNF^{flox/+}; TrkB^{GFP/+}$ control mice, and (B) $K5^{Cre/+}; BDNF^{flox/flox}; TrkB^{GFP/+}$ mutant mice; note the selective tuning in the opposite, caudal direction in the latter.

(C) Examples of responses to controlled deflection of three different single hairs in the receptive fields of A δ -LTMRs in $K5^{Cre/+}; BDNF^{flox/flox}; TrkB^{GFP/+}$ (mutant) mice.

(D) Vector plot of average responses to multiple deflections of a single hair in the four cardinal directions and examples of different vector plots elicited by movements of single hairs in the receptive fields of A δ -LTMRs in $K5^{Cre/+}; BDNF^{flox/flox}; TrkB^{GFP/+}$ mice. Note the random, whorled orientations compared to controls (Figure 1).

(E) Shown are the means \pm SD of spike numbers obtained from recordings of six neurons of each genotype, with responses to deflections of hairs in 60 and 45 individual locations from mutant and wild-type mice, respectively. Cumulative vector plot of all individual locations analyzed throughout the receptive fields of A δ -LTMRs in $K5^{Cre/+}; BDNF^{flox/flox}; TrkB^{GFP/+}$ mice (orange); note that the average response of all hairs combined lacks the direction-selective tuning observed overall in A δ -LTMRs from control mice (blue). The latter represent replotted data from the ex vivo recording results from control mice, shown in Figure 1H.

See also Figure S6.

small, \sim 80–90 nm protrusion extending between TSC processes, and these lanceolate ending protrusions lie within close proximity to the hair follicle epithelial cell basal lamina (Li and Ginty, 2014). Moreover, fine filament-like structures emanate from hemi-desmosomes positioned along the outer membranes of hair follicle epithelial cells, and these filaments, or putative tethers, extend through the basal lamina and appear to come in direct contact with the plasma membranes of both LTMR lanceolate axonal ending protrusions and TSCs (Li and Ginty, 2014). Because *in vitro* findings implicate tethers emanating from primary mechanosensory neurons as essential for mechanotransduction (Chiang et al., 2011), we previously speculated that the filamentous connections between epithelial cells and LTMR membranes mediate mechanotransduction *in vivo*, trans-

ducing hair deflection into lanceolate axon depolarization and LTMR excitation. Such a function of putative lanceolate complex tethers may thus be analogous to that of the tip links that extend between stereocilia of mechanosensory hair cells of the inner ear and that mediate mechanotransduction in the auditory and vestibular systems. If the LTMR-hair follicle tether model is indeed correct, then deflection of hairs in the R direction would be expected to pull on most putative tethers connecting hair follicle epithelial cells and A δ -LTMR lanceolate axon membranes, leading to excitation of the A δ -LTMR. Hair deflection in the C direction, on the other hand, would be expected to relax most putative tethers thus failing to open mechanically sensitive channels in A δ -LTMR lanceolate axonal membranes. As such, enrichment of lanceolate complexes on the caudal side of hair follicles

would underlie a greater sensitivity to hair deflection in the R direction than in the C direction. Identification of the protein composition of the putative epithelial cell-to-LTMR tether would enable ablation experiments that could test this and related ideas for hair deflection-LTMR mechanotransduction and thus provide understanding of the mechanism of A δ -LTMR direction selectivity.

In addition to the somatosensory system, direction selectivity is a key feature of other sensory systems, notably the auditory and visual systems. In the visual system, DS-RGCs report on the movement of images across their receptive fields. A morphological basis of direction selectivity for at least three DS-RGC subtypes was revealed by the discovery that their dendrites are polarized in the orientation of their preferred direction of movement of visual stimuli (Kim et al., 2008; Vaney et al., 2012). Thus, at least some DS-RGC subtypes and, as described here, A δ -LTMRs have a morphological basis for their direction-selective tuning properties. In another noteworthy parallel, the central representations of peripheral receptive fields are topographically organized in both the visual and somatosensory systems. Retinotopic and somatotopic organization of the central representations of peripheral receptive fields may underlie direction selectivity in these sensory systems by virtue of temporal contrasts between the activities of peripheral neurons whose receptive fields are adjacent or in close proximity to one another. And yet, as reported here for the somatosensory system and previously for the visual system, both systems also have individual peripheral components, A δ -LTMRs and DS-RGS, that are themselves tuned to the direction of stimulus movement. Are topographic maps and direction-selective peripheral units both employed for the central representation and interpretation of direction selectivity? Retinotopy- and somatotopy-based mechanisms for computing the direction of stimulus movement would necessarily rely on temporal contrasts of the spiking of two or more neurons with nonoverlapping receptive fields. On the other hand, in the somatosensory system, we find that A δ -LTMRs can encode information about the direction of movement of a single hair. Thus, somatotopy-based computations and the direction-selective information coded by individual A δ -LTMRs are distinct and may play complementary roles in perception of the direction of stimulus movement across hairy skin. A δ -LTMRs may enhance the contrast of somatotopy-based computations that are driven by neurons that need not be direction-selective themselves, such as A β RA-LTMRs. Additionally, A δ -LTMRs may report on the direction of hair deflection when all hairs on a skin area are simultaneously stimulated in the same direction, for example during exposure to a gentle breeze. In this scenario, all A δ -LTMRs innervating the stimulated area are expected to respond in a similar manner. For this particular example, a somatotopy-based mechanism would lack temporal contrast between individual neuronal responses and therefore may be ineffective in reporting directionality of the stimulus. On the other hand, individually tuned, direction-selective A δ -LTMRs are predicted to report on directionality independent of contrast between neurons having adjacent receptive fields and may therefore contribute to the perception of direction-selective

movement of hairs under such a condition. To test these and related ideas, it will be important to establish the relative contributions of A δ -LTMR direction selectivity to the perception of hair deflection direction and object movement across hairy skin.

How is direction-selective information, extracted from the skin by A δ -LTMRs, conveyed to the brain? Insights into this question may be gleaned from studies of other sensory systems, and we again turn to direction-selective circuits of the visual system for analogy. DS-RGCs and nondirection-selective RGCs project to distinct regions of the thalamus, where higher order neurons then project to distinct layers of visual cortex (Cruz-Martín et al., 2014). Thus, processing of direction-selective visual information and other visual information, such as light contrast, occurs at least in part via distinct brain circuitries. In the somatosensory system, many neurons of primate somatosensory cortex, representing both glabrous and hairy skin, are tuned to the direction of stimulus movement across the skin (Costanzo and Gardner, 1980; Hyvärinen and Poranen, 1978; Whitsel et al., 1972, 1978). Moreover, perception of the direction of stimulus movement across the skin requires the integrity of the dorsal column pathway (Bender et al., 1982; Vierck, 1974), a major ascending tract that contains primary branches of A β -LTMRs and postsynaptic (indirect) dorsal column pathway neurons, both of which terminate in the dorsal column nuclei of the brainstem, the gracile and cuneate nuclei. Indeed, while disruption of the dorsal column eliminates the perception of direction of tactile stimulus movement in primates, dorsal column lesions alone have little or no impact on the general detection of stimulus motion. Lesions of both the dorsal columns and the dorso-lateral funiculus disrupt perception of both stimulus direction and motion (Vierck, 1974; Wall and Noordenbos, 1977). We find that A δ -LTMR central projections terminate in lamina II-III of the dorsal horn, however, unlike A β -LTMR subtypes, A δ -LTMRs do not have a branch that ascends the dorsal column. Therefore, we speculate that direction-selective information about hair deflection, at least that which is extracted by A δ -LTMRs, is processed in the spinal cord dorsal horn and subsequently conveyed to the brain via postsynaptic dorsal column neurons comprising the indirect dorsal column pathway. The identity of postsynaptic partners of A δ -LTMRs in the dorsal horn, how A δ -LTMR direction-selective information is conveyed from the dorsal horn to the brain, and the relative contributions of A δ -LTMRs to direction-selective tuning properties of neurons in the neocortex and to the perception of direction of object movement across hairy skin await findings of future interrogation of these fascinating neurons.

EXPERIMENTAL PROCEDURES

Mouse Lines

The *TrkB^{tauEGFP}* (Li et al., 2011), *TH^{CreER}* (Badea et al., 2009), *Ret^{CreER}* (Luo et al., 2009), *T-Cre* (Perantoni et al., 2005), *Wnt1-Cre* (Danielian et al., 1998), *K5-Cre* (Ramirez et al., 2004), *Rosa26-TdTomato* (strain Ai9; Jackson Laboratory), *Rosa26-iAP* (Badea et al., 2009), *BDNF-loxp* and *BDNF-LacZ* (Gorski et al., 2003), *NT3-LacZ* (Fariñas et al., 1994), and *NT4-LacZ* (EUCOMM) mouse lines have been described. *Split^{Cre}* mice express Cre recombinase in A β RA-LTMRs and will be described elsewhere. *TrkB^{CreER}* mouse line generation is described in Extended Experimental Procedures.

Electrophysiological Recordings

Intracellular electrophysiological recordings from TrkB^{CreER}-labeled skin sensory neurons were obtained in adult animals, using either ex vivo somatosensory system preparations and DRG neurons innervating dorsal back skin or in vivo preparations and trigeminal ganglion neurons innervating the face. Generation of the ex vivo cutaneous somatosensory system preparation used in the present studies has been described in detail (Li et al., 2011; Woodbury et al., 2001). Generation of in vivo adult mouse preparation was modified from procedures detailed elsewhere (Boada and Woodbury, 2007), as described in Extended Experimental Procedures.

Receptive Field Analyses

To investigate the response properties of LTMRs to directional movement of hairs, fine-grained receptive field (RF) analyses were conducted by characterizing the sensitivity of cells to controlled movements of hairs located in multiple spots throughout the RF; in many cases this amounted to systematic mapping of the RF in a hair-by-hair manner. To achieve controlled directional movement of hairs, a variety of customized probes were tested before settling on a forked, comb-like device that allowed us to unambiguously isolate and trap individual hairs under direct visual observation at high magnification. This probe consisted of two 0.1 mm diameter minuten pins glued together in parallel, leaving ~0.5 mm of the tapered tips exposed, the latter bent at ~90° to form a miniaturized two-tined rake (see Figure S2F). This rake was oriented orthogonally to the skin and against the grain using a manual micromanipulator equipped with precision lead screws (100 TPI). Hairs normally laid flat, and once trapped, erected to an ~45° angle prior to initiating a series of alternating movements in the rostrocaudal (RC) and dorsoventral (DV) planes. These movements were repeated at ~1 Hz and displaced the probe ~0.5 mm at a rate of ~2 mm/s. To control for the possibility that responses in the cell might reflect slippage of the hair shaft in the yoke of the probe during movement (i.e., sensitivity to vibration produced by cuticular irregularities along the hair shaft), we used a different probe tip in some experiments that was fashioned from a single minuten pin coated in glue from sticky mouse traps (Stick-Em, JT Eaton); this glue-tipped probe could be affixed securely to individual hairs to dampen or completely prevent potential microvibrational influences during movement. The numbers of spikes elicited by movements in each direction were counted and averaged across each spot tested in the receptive field; responses to movements in different directions were compared using either Student's *t* tests and/or ANOVAs with Tukey's post hoc corrections (Origin Pro 8).

Histological Analyses

Immunohistochemistry of tissue sections, whole mount immunohistochemistry, in situ hybridization, whole mount PLAP staining of the skin and spinal cord, and LacZ staining were done using standard procedures (see Extended Experimental Procedures for details).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and six figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.038>.

AUTHOR CONTRIBUTIONS

M.R., C.Y.H., and V.E.A. contributed equally to this study. M.R. generated the TrkB-CreER mouse line with the help of Dori Reimert and the Johns Hopkins Transgenic Facility. M.R. and V.E.A. characterized the line. M.R., C.Y.H., and V.E.A. did the developmental and morphological analyses of Aδ-LTMRs.

ACKNOWLEDGMENTS

We thank members of the D.D.G. laboratory for discussions and comments on this study, Dori Reimert for help with generation of the TrkB-CreER mouse line, Kevin Jones and Jeremy Nathans for providing mouse lines, Danilo Boada for help with development of trigeminal preparations, and the Johns Hopkins Trans-

genic Facility for blastocyst injections. This work was supported by NIH grants F32 NS077836 (V.E.A.), NS44094 (C.J.W.), NS34814 (D.D.G.), and DE022750 (D.D.G.), and the Johns Hopkins NINDS Core Imaging Facility MH084020 (D.D.G.). D.D.G. is an investigator of the Howard Hughes Medical Institute.

Received: October 18, 2014

Revised: November 11, 2014

Accepted: November 19, 2014

Published: December 18, 2014

REFERENCES

- Adrian, E.D., and Zotterman, Y. (1926). The impulses produced by sensory nerve endings: Part 3. Impulses set up by Touch and Pressure. *J. Physiol.* 61, 465–483.
- Badea, T.C., Hua, Z.L., Smallwood, P.M., Williams, J., Rotolo, T., Ye, X., and Nathans, J. (2009). New mouse lines for the analysis of neuronal morphology using CreER(T)/loxP-directed sparse labeling. *PLoS ONE* 4, e7859.
- Bender, M.B., Stacy, C., and Cohen, J. (1982). Agrophesthesia. A disorder of directional cutaneous kinesthesia or a disorientation in cutaneous space. *J. Neurol. Sci.* 53, 531–555.
- Boada, M.D., and Woodbury, C.J. (2007). Physiological properties of mouse skin sensory neurons recorded intracellularly in vivo: temperature effects on somal membrane properties. *J. Neurophysiol.* 98, 668–680.
- Brown, A.G., and Iggo, A. (1967). A quantitative study of cutaneous receptors and afferent fibres in the cat and rabbit. *J. Physiol.* 193, 707–733.
- Burgess, P.R., Petit, D., and Warren, R.M. (1968). Receptor types in cat hairy skin supplied by myelinated fibers. *J. Neurophysiol.* 31, 833–848.
- Chiang, L.Y., Poole, K., Oliveira, B.E., Duarte, N., Sierra, Y.A., Bruckner-Tuderman, L., Koch, M., Hu, J., and Lewin, G.R. (2011). Laminin-332 coordinates mechanotransduction and growth cone bifurcation in sensory neurons. *Nat. Neurosci.* 14, 993–1000.
- Costanzo, R.M., and Gardner, E.P. (1980). A quantitative analysis of responses of direction-sensitive neurons in somatosensory cortex of awake monkeys. *J. Neurophysiol.* 43, 1319–1341.
- Cruz-Martin, A., El-Danaf, R.N., Osakada, F., Sriram, B., Dhande, O.S., Nguyen, P.L., Callaway, E.M., Ghosh, A., and Huberman, A.D. (2014). A dedicated circuit links direction-selective retinal ganglion cells to the primary visual cortex. *Nature* 507, 358–361.
- Danielian, P.S., Muccino, D., Rowitch, D.H., Michael, S.K., and McMahon, A.P. (1998). Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Current biology: CB* 8, 1323–1326.
- Essick, G.K., and Edin, B.B. (1995). Receptor encoding of moving tactile stimuli in humans. II. The mean response of individual low-threshold mechanoreceptors to motion across the receptive field. *The Journal of neuroscience* 15, 848–864.
- Essick, G.K., and Whitsel, B.L. (1985). Factors influencing cutaneous directional sensitivity: a correlative psychophysical and neurophysiological investigation. *Brain Res.* 357, 213–230.
- Fariñas, I., Jones, K.R., Backus, C., Wang, X.Y., and Reichardt, L.F. (1994). Severe sensory and sympathetic deficits in mice lacking neurotrophin-3. *Nature* 369, 658–661.
- Gorski, J.A., Zeiler, S.R., Tamowski, S., and Jones, K.R. (2003). Brain-derived neurotrophic factor is required for the maintenance of cortical dendrites. *The Journal of neuroscience* 23, 6856–6865.
- Gottschaldt, K.M., and Vahle-Hinz, C. (1981). Merkel cell receptors: structure and transducer function. *Science* 214, 183–186.
- Greenspan, J.D. (1992). Influence of velocity and direction of surface-parallel cutaneous stimuli on responses of mechanoreceptors in feline hairy skin. *J. Neurophysiol.* 68, 876–889.
- Grothe, B., Pecka, M., and McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiol. Rev.* 90, 983–1012.

- Hamann, W. (1995). Mammalian cutaneous mechanoreceptors. *Prog. Biophys. Mol. Biol.* 64, 81–104.
- Horch, K.W., Tuckett, R.P., and Burgess, P.R. (1977). A key to the classification of cutaneous mechanoreceptors. *J. Invest. Dermatol.* 69, 75–82.
- Hyvärinen, J., and Poranen, A. (1978). Movement-sensitive and direction and orientation-selective cutaneous receptive fields in the hand area of the post-central gyrus in monkeys. *J. Physiol.* 283, 523–537.
- Kim, I.J., Zhang, Y., Yamagata, M., Meister, M., and Sanes, J.R. (2008). Molecular identification of a retinal cell type that responds to upward motion. *Nature* 452, 478–482.
- Klein, R., Nanduri, V., Jing, S.A., Lamballe, F., Tapley, P., Bryant, S., Cordon-Cardo, C., Jones, K.R., Reichardt, L.F., and Barbacid, M. (1991). The *trkB* tyrosine protein kinase is a receptor for brain-derived neurotrophic factor and neurotrophin-3. *Cell* 66, 395–403.
- Klein, R., Lamballe, F., Bryant, S., and Barbacid, M. (1992). The *trkB* tyrosine protein kinase is a receptor for neurotrophin-4. *Neuron* 8, 947–956.
- Koltzenburg, M., Stucky, C.L., and Lewin, G.R. (1997). Receptive properties of mouse sensory neurons innervating hairy skin. *J. Neurophysiol.* 78, 1841–1850.
- Kwegyir-Aful, E.E., Marella, S., and Simons, D.J. (2008). Response properties of mouse trigeminal ganglion neurons. *Somatosens. Mot. Res.* 25, 209–221.
- Li, L., and Ginty, D.D. (2014). The structure and organization of lanceolate mechanosensory complexes at mouse hair follicles. *eLife* 3, e01901.
- Li, L., Rutlin, M., Abaira, V.E., Cassidy, C., Kus, L., Gong, S., Jankowski, M.P., Luo, W., Heintz, N., Koerber, H.R., et al. (2011). The functional organization of cutaneous low-threshold mechanosensory neurons. *Cell* 147, 1615–1627.
- Lichtenstein, S.H., Carvell, G.E., and Simons, D.J. (1990). Responses of rat trigeminal ganglion neurons to movements of vibrissae in different directions. *Somatosens. Mot. Res.* 7, 47–65.
- Liu, Y., Rutlin, M., Huang, S., Barrick, C.A., Wang, F., Jones, K.R., Tessarollo, L., and Ginty, D.D. (2012). Sexually dimorphic BDNF signaling directs sensory innervation of the mammary gland. *Science* 338, 1357–1360.
- Luo, W., Enomoto, H., Rice, F.L., Milbrandt, J., and Ginty, D.D. (2009). Molecular identification of rapidly adapting mechanoreceptors and their developmental dependence on ret signaling. *Neuron* 64, 841–856.
- Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* 13, 133–140.
- Maruhashi, J., Mizuguchi, K., and Tasaki, I. (1952). Action currents in single afferent nerve fibres elicited by stimulation of the skin of the toad and the cat. *J. Physiol.* 117, 129–151.
- Olausson, H., Lamarque, Y., Backlund, H., Morin, C., Wallin, B.G., Starck, G., Ekholm, S., Strigo, I., Worsley, K., Vallbo, A.B., and Bushnell, M.C. (2002). Unmyelinated tactile afferents signal touch and project to insular cortex. *Nat. Neurosci.* 5, 900–904.
- Perantoni, A.O., Timofeeva, O., Naillat, F., Richman, C., Pajni-Underwood, S., Wilson, C., Vainio, S., Dove, L.F., and Lewandoski, M. (2005). Inactivation of *FGF8* in early mesoderm reveals an essential role in kidney development. *Development* 132, 3859–3871.
- Ramirez, A., Page, A., Gandarillas, A., Zanet, J., Pibre, S., Vidal, M., Tusell, L., Genesca, A., Whitaker, D.A., Melton, D.W., and Jorcano, J.L. (2004). A keratin K5Cre transgenic line appropriate for tissue-specific or generalized Cre-mediated recombination. *Genesis* 39, 52–57.
- Ray, R.H., Mallach, L.E., and Kruger, L. (1985). The response of single guard and down hair mechanoreceptors to moving air-jet stimulation. *Brain Res.* 346, 333–347.
- Trenholm, S., Johnson, K., Li, X., Smith, R.G., and Awatramani, G.B. (2011). Parallel mechanisms encode direction in the retina. *Neuron* 71, 683–694.
- Tuckett, R.P. (1978). Response of cutaneous hair and field mechanoreceptors in cat to paired mechanical stimuli. *J. Neurophysiol.* 41, 150–156.
- Vaney, D.I., Sivy, B., and Taylor, W.R. (2012). Direction selectivity in the retina: symmetry and asymmetry in structure and function. *Nat. Rev. Neurosci.* 13, 194–208.
- Vierck, C.J., Jr. (1974). Tactile movement detection and discrimination following dorsal column lesions in monkeys. *Exp. Brain Res.* 20, 331–346.
- Wall, P.D., and Noordenbos, W. (1977). Sensory functions which remain in man after complete transection of dorsal columns. *Brain* 100, 641–653.
- Wei, W., and Feller, M.B. (2011). Organization and development of direction-selective circuits in the retina. *Trends Neurosci.* 34, 638–645.
- Whitsel, B.L., Roppolo, J.R., and Werner, G. (1972). Cortical information processing of stimulus motion on primate skin. *J. Neurophysiol.* 35, 691–717.
- Whitsel, B.L., Dreyer, D.A., and Hollins, M. (1978). Representation of moving stimuli by somatosensory neurons. *Fed. Proc.* 37, 2223–2227.
- Woodbury, C.J., and Koerber, H.R. (2007). Central and peripheral anatomy of slowly adapting type I low-threshold mechanoreceptors innervating trunk skin of neonatal mice. *J. Comp. Neurol.* 505, 547–561.
- Woodbury, C.J., Ritter, A.M., and Koerber, H.R. (2001). Central anatomy of individual rapidly adapting low-threshold mechanoreceptors innervating the “hairy” skin of newborn mice: early maturation of hair follicle afferents. *J. Comp. Neurol.* 436, 304–323.
- Zotterman, Y. (1939). Touch, pain and tickling: an electro-physiological investigation on cutaneous sensory nerves. *J. Physiol.* 95, 1–28.

Detecting Envelope Stress by Monitoring β -Barrel Assembly

Seung-Hyun Cho,^{1,2,4} Joanna Szewczyk,^{1,2,4} Christina Pesavento,^{3,4} Matylda Zietek,³ Manuel Banzhaf,³ Paula Roszczenko,^{1,2} Abir Asmar,^{1,2} Géraldine Laloux,² Ann-Kristin Hov,³ Pauline Leverrier,² Charles Van der Henst,^{1,2} Didier Vertommen,² Athanasios Typas,^{3,*} and Jean-François Collet^{1,2,*}

¹WELBIO

²de Duve Institute

Université catholique de Louvain, Avenue Hippocrate 75, Brussels 1200, Belgium

³European Molecular Biology Laboratory, Genome Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany

⁴Co-first authors

*Correspondence: typas@embl.de (A.T.), jfcollet@uclouvain.be (J.-F.C.)

<http://dx.doi.org/10.1016/j.cell.2014.11.045>

SUMMARY

The cell envelope protects bacteria from their surroundings. Defects in its integrity or assembly are sensed by signal transduction systems, allowing cells to rapidly adjust. The Rcs phosphorelay responds to outer membrane (OM)- and peptidoglycan-related stress in enterobacteria. We elucidated how the OM lipoprotein RcsF, the upstream Rcs component, senses envelope stress and activates the signaling cascade. RcsF interacts with BamA, the major component of the β -barrel assembly machinery. In growing cells, BamA continuously funnels RcsF through the β -barrel OmpA, displaying RcsF on the cell surface. This process spatially separates RcsF from the downstream Rcs component, which we show is the inner membrane protein IgaA. The Rcs system is activated when BamA fails to bind RcsF and funnel it to OmpA. Newly synthesized RcsF then remains periplasmic, interacting with IgaA to activate the cascade. Thus RcsF senses envelope damage by monitoring the activity of the Bam machinery.

INTRODUCTION

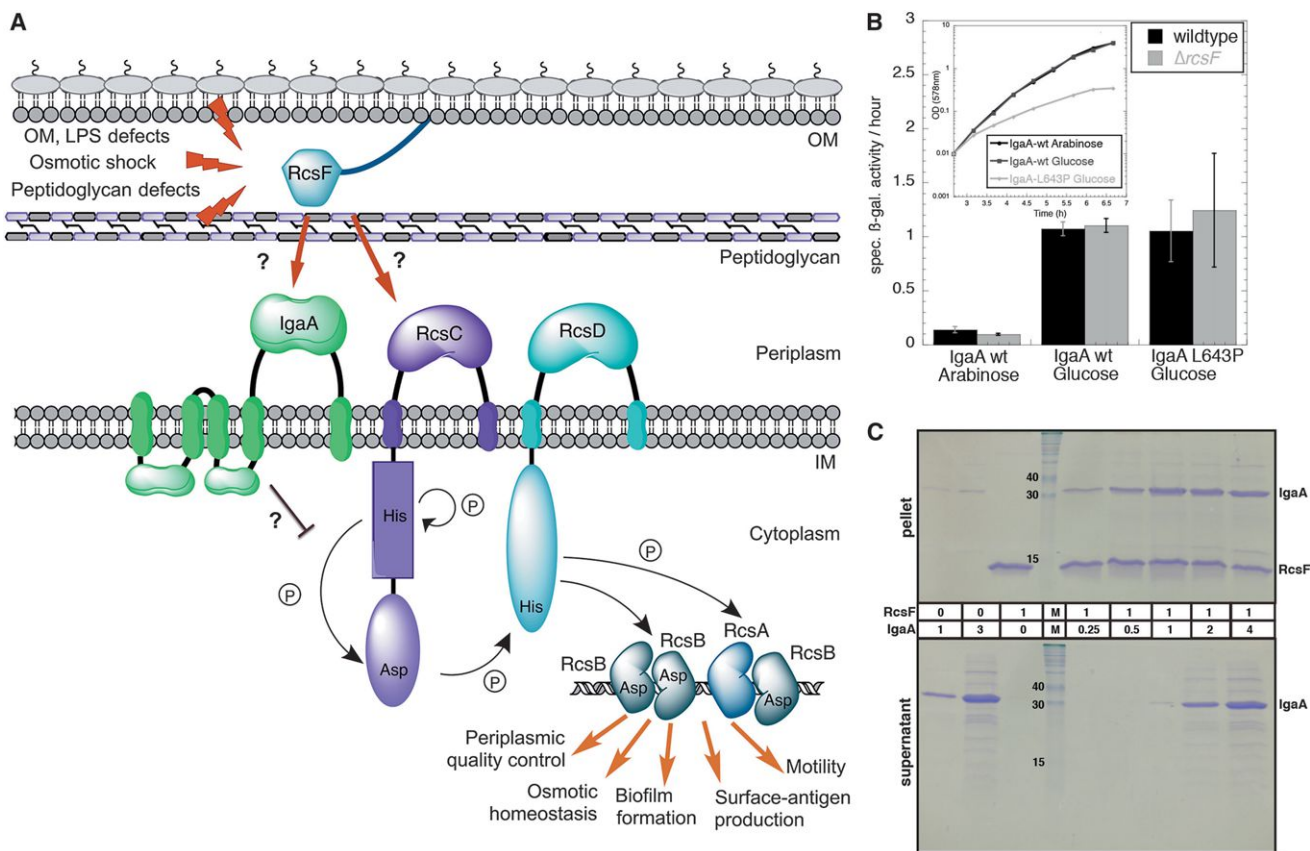
The cell envelope of Gram-negative bacteria consists of two membranes, separated by a viscous periplasm that contains the peptidoglycan (PG). The envelope is a permeability and structural barrier, which is essential for cell shape and growth, and serves as interface to the environment. To monitor their surroundings, bacteria use a number of signaling cascades that transduce the information from their envelope to their decision center (cytoplasm).

Bacteria also constantly monitor the growth and assembly status of their envelope. As this compartment is devoid of energy, transport and assembly of its structural subunits are controlled by multicomponent protein machineries, which span the envelope and utilize energy from the cytoplasm (Silhavy et al., 2010). These machineries must tightly coordinate their function, as assembly of the different envelope layers is interlinked (com-

ponents synthesizing one layer often reside in or interact with the other layer [Typas et al., 2012]) and coupled to growth rate. It is therefore vital for bacteria to detect when envelope assembly is perturbed and to rapidly fix or contain the damage.

Many signaling systems can sense envelope perturbations in *E. coli* and mount a repair and/or a preventive response to minimize the damage. The best understood system is the σ^E stress response, which uses three proteins to sense accumulation of unassembled OM porins (OMPs) and LPS in the periplasm (Lima et al., 2013; Walsh et al., 2003). In response to these insults, σ^E directs the transcription of genes that facilitate OMP and LPS assembly, transport, and turnover (Rhodius et al., 2006) and of small RNAs that block the expression of various OMPs and of the most abundant OM lipoprotein, Lpp (Gogol et al., 2011; Guo et al., 2014). In contrast to σ^E , other signal transduction systems associated with envelope damage surveillance in *E. coli* are less well understood, and in most cases, the direct activating signal is obscure.

The Rcs phosphorelay is one of the most complex bacterial signaling systems. It is induced mostly by OM and cell wall damage (Evans et al., 2013; Farris et al., 2010; Laubacher and Ades, 2008; Majdalani and Gottesman, 2005). In response to these cues, Rcs controls the expression of genes involved in motility, biofilm formation, virulence, and periplasmic quality control (Majdalani and Gottesman, 2005). The complexity of the system, both at the input and output level, has been an obstacle in dissecting it and fully addressing its physiological role. Unlike typical two-component systems consisting of an inner membrane (IM) sensor histidine kinase (HK) and a cytoplasmic response regulator (RR), the Rcs system has at least six components (Figure 1A). In addition to RcsC (HK) and RcsB (RR), the system contains an intermediate IM phosphorelay protein, RcsD, an auxiliary nonphosphorylatable transcription factor RcsA, and two proteins that act upstream of the phosphorelay cascade and are associated with signal sensing, YrfF and RcsF (Cano et al., 2002; Castanié-Cornet et al., 2006). YrfF is an IM protein, mostly characterized in *Salmonella* Typhimurium, which downregulates the Rcs pathway by an unknown mechanism (Domínguez-Bernal et al., 2004). Deletion of the gene is lethal, unless the Rcs phosphorelay is also inactivated (Cano et al., 2002). *yrfF* has been renamed to *igaA* in *S. Typhimurium* and we use the same nomenclature for the *E. coli* gene in this paper.



that *igaA* is also essential in *E. coli* and using an established pipeline for large-scale testing of genetic interactions (Typas et al., 2008), we showed that, among a collection of knockout mutants for all *E. coli* nonessential genes (Baba et al., 2006), an *igaA* deletion was viable only when combined with deletions of *rcsB*, *rcsC*, and *rcsD* (Figure S1A available online). Thus, *E. coli* and *S. Typhimurium* *igaA* play similar roles. Importantly, deletion of *rcsF* did not suppress *igaA* lethality, implying that IgaA lies downstream of RcsF in the signaling cascade. In agreement with this configuration, depletion of *igaA* activated the Rcs system independently of RcsF (Figures 1B and S1B).

We next tested whether IgaA and RcsF are physically linked by expressing a tagged version of the only periplasmic domain of IgaA (IgaA_{peri}; ~32 kDa) in the periplasm and pulling down its interaction partners after crosslinking with DTSSP (3,3'-dithiobis[sulfosuccinimidylpropionate]), which cannot cross the IM. RcsF was identified by both mass spectrometry (MS) and western blot (Figure S1C and Table S1). Likewise, a purified tag-less version of IgaA_{peri} and a soluble His-tagged version of RcsF directly interacted, forming a complex with a 1:1 stoichiometry (Figures 1C and S1D). These results support a model in which RcsF activates the Rcs system by interacting with IgaA, likely alleviating its inhibitory effect on the signaling cascade.

By Forming a Complex with the β -Barrel Proteins OmpA and BamA, RcsF Is Occluded from IgaA

Wild-type RcsF turns on the signaling cascade only upon envelope stress, suggesting that RcsF is physically occluded from IM IgaA under steady-state growth. This physical occlusion is tightly interconnected with the OM location of RcsF, as rerouting RcsF to the IM (RcsF_{IM}) or expressing it as a soluble periplasmic protein (RcsF_{peri}), constitutively activates the Rcs system (Farris et al., 2010; Tao et al., 2012). We therefore looked for the underlying occlusion mechanism.

RcsF is composed of a 31-residue intrinsically disordered N-terminal linker (S17-T47), which connects its globular domain (P48-K134; referred to as "signaling domain") to the lipidated cysteine residue anchoring the protein to the OM (Figure S2A) (Leverrier et al., 2011). We first tested whether this linker is cleaved under stress, releasing RcsF in the periplasm, by fractionating cells exposed to various Rcs-inducing stresses. RcsF was never detected in the soluble fraction (Figure S2B). We then determined whether RcsF was occluded from IgaA by being sequestered by other proteins. To find proteins interacting with RcsF, we performed in vivo DTSSP crosslinking in both $\Delta rcsF$ and wild-type cells. Three RcsF-containing protein complexes were detected (Figure 2A; marked as 1, 2, and 3). To identify them, the RcsF-interacting complexes were immunoprecipitated and analyzed by MS after reversing the crosslinks. We identified BamA, the core protein for β -barrel assembly, and the β -barrels OmpA, OmpC and OmpF as potential RcsF interacting partners (Table S2A). We further verified these interactions by analyzing the immunoprecipitated samples by western blot using antibodies specific for the interacting proteins. We confirmed that the ~115 kDa band (complex 1) contained BamA (~100 kDa) (Figure S2C). Only OmpA (38 kDa) could be detected in the ~55 kDa complex (complex 2) (Figure S2C), but not OmpC (40 kDa) and OmpF (39 kDa) (data not shown),

suggesting that OmpA was the major interaction partner involved in this complex. Consistently, upon depleting the essential BamA or deleting *ompA*, the respective complexes disappeared (Figure 2B). Using an *lpp* deletion mutant, we identified that the protein involved in complex 3 (~25 kDa) was Lpp (8 kDa), the most abundant OM lipoprotein (Figure S2D).

We further verified the specificity of these interactions by site-specific photocrosslinking, inserting the crosslinkable amino acid *p*-benzoyl-L-phenylalanine (pBpa) at 25 specific positions in RcsF. Thereby we could map the interaction interface of RcsF with its binding partners more precisely, and lower the risk of nonspecific interactions (pBpa can only form covalent bonds with residues at a very close proximity [3 Å], whereas DTSSP has a 12 Å spacer). We selected 21 residues located on the surface of the signaling domain and 4 located in the N-terminal linker (Figure S3A). Following UV-exposure, 6/25 variants formed the previously observed 55 kDa complex with OmpA (Figures 2C and S3B). The identity of OmpA was confirmed by MS after immunoprecipitating RcsF_{K40pBPA}-OmpA (OmpC and OmpF were not detected in the sample, Table S2B). High complex levels were observed when pBpa was inserted in the N-terminal linker and at the tip of the signaling domain of RcsF (Figure S3C). Four of these variants could also form the 115 kDa complex corresponding to BamA-RcsF (Figures 2C and S3D), which was confirmed with a BamA antibody (Figure S3D). As none of the 25 pBpa-containing variants was found in complex with Lpp, and as Lpp is the most abundant protein in *E. coli*, which could lead to nonspecific interactions, we decided not to follow up on this interaction. Altogether, these results indicated that RcsF interacts specifically with BamA and OmpA. Importantly, the levels of the RcsF_{K40pBPA}-OmpA complex were ~30%–40% of those of free RcsF, indicating that 25%–30% of total RcsF is bound to OmpA (Figures S3E and S3F). Given that photocrosslinking efficiency at optimal conditions can reach 40% (Zhang et al., 2011), we concluded that most RcsF is in complex with OmpA.

The Bam Machinery Assembles the RcsF-OmpA Complex and Is Key for the Sensing Role of RcsF

The interactions of RcsF with BamA and OmpA suggested that in nonstress conditions RcsF is occluded from IgaA by interacting with OM proteins, but that these interactions are disturbed upon envelope stress, enabling RcsF to interact with IgaA. If OmpA and BamA occlude RcsF from IgaA under nonstress conditions, then the Rcs system should be activated when *bamA* or *ompA* are knocked down/out. We found that an *ompA* deletion induced the Rcs system by ~3-fold, with induction being dependent on RcsF (Figure 2D). As OmpC and OmpF were also identified as RcsF partners, we tested the effect of deleting *ompC* or *ompF* on Rcs activity. Whereas the system was only marginally induced in the *ompC* mutant, the *ompF* deletion had no impact (Figure 2D). When *ompC* and *ompF* deletions were combined together or with *ompA*, synergistic effects were observed, but the absence of OmpA was clearly the most important contributor of the three to the activation of Rcs (Figure 2D), consistently with our interaction data (Figures 2 and S3). In contrast to the *omp* mutants, the Rcs system was fully induced in the *bamA* knockdown (*bamA101*) mutant (Figure 2D). In this strain, BamA levels decrease ~5-fold without

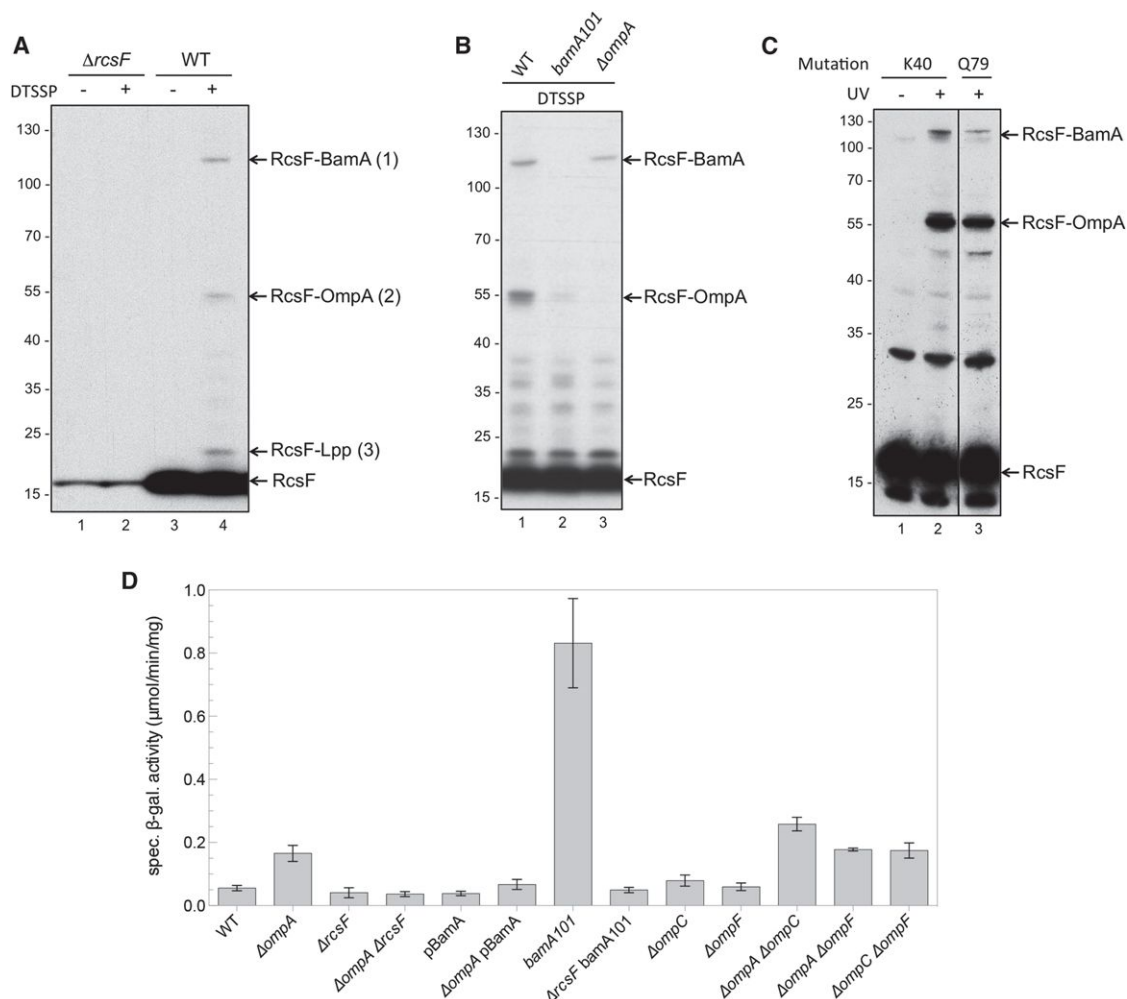


Figure 2. RcsF Forms Complexes with BamA and OmpA In Vivo, Which Prevents It from Activating the Signaling Cascade

(A and B) In vivo chemical crosslinking of RcsF in the periplasm. Wild-type and $\Delta rcsF$ cells were harvested at mid-log phase, washed and incubated with or without 1 mM DTSSP for 30 min. The reaction was quenched by addition of glycine (0.1 M), proteins were isolated by TCA precipitation, resuspended in sample buffer (without DTT) and subjected to SDS-PAGE and immunoblot analysis with an anti-RcsF antibody. Three complexes were observed (A), which were identified as RcsF-BamA (1), RcsF-OmpA (2) and RcsF-Lpp (3) (Figures S2C and S2D). Complexes 1 and 2 disappeared when repeating the DTSSP crosslinking in $bamA101$ and $\Delta ompA$ cells, respectively (B).

(C) In vivo site-specific photocrosslinking of RcsF. Cells expressing RcsF(K40pBPA)-Flag-His or RcsF(Q79pBPA)-Flag-His from low-copy plasmids were irradiated with UV light (lanes 2 and 3) or not (lane 1), and protein samples were subjected to immunoblot analysis with an anti-RcsF antibody. BamA and OmpA were crosslinked with both RcsF mutants.

(D) $ompA$ deletion and BamA depletion activate the Rcs system. An $ompA$ deletion and a $bamA$ knockdown ($bamA101$) activated the Rcs system only in the presence of RcsF. Overexpression of $bamA$ could restore basal Rcs activity in the $\Delta ompA$ mutant. Deletions of $ompC$ or $ompF$ had marginal or no effects on Rcs activity. Double omp mutants induced the Rcs system further with OmpA being the most contributing factor. A chromosomal $rprA::lacZ$ fusion was used to monitor Rcs activity, and specific β -galactosidase (β -gal) activity was measured from cells at mid-log phase ($OD_{578} = 0.2$ – 1). Error bars depict standard deviations ($n > 4$). See also Figures S2 and S3.

significantly compromising β -barrel assembly (Aoki et al., 2008). These results are in agreement with the idea that OMPs (mainly OmpA) and BamA occlude RcsF from IgaA and suggest a dominant role for BamA in this process.

To dissect the signaling system further, we examined RcsF-BamA and RcsF-OmpA complex formation during Rcs activation by polymyxin B, A22, or mecillinam. All three chemicals induce the Rcs by targeting different cellular structures but always in an RcsF-dependent manner (Figure S4A) and without signifi-

cantly affecting the transport of RcsF to the OM (Figure S4B). The cationic antimicrobial peptide polymyxin B damages the OM by perturbing the LPS leaflet, A22 inhibits the actin-like MreB, and the β -lactam antibiotic mecillinam inhibits the essential transpeptidase PBP2. After addition of subinhibitory amounts of each drug, we observed a sharp decrease in the levels of the BamA-RcsF complex within the timeframe that the Rcs system would be activated, while OmpA-RcsF remained largely unaffected (Figures 3A–3C). We also probed a *galU*

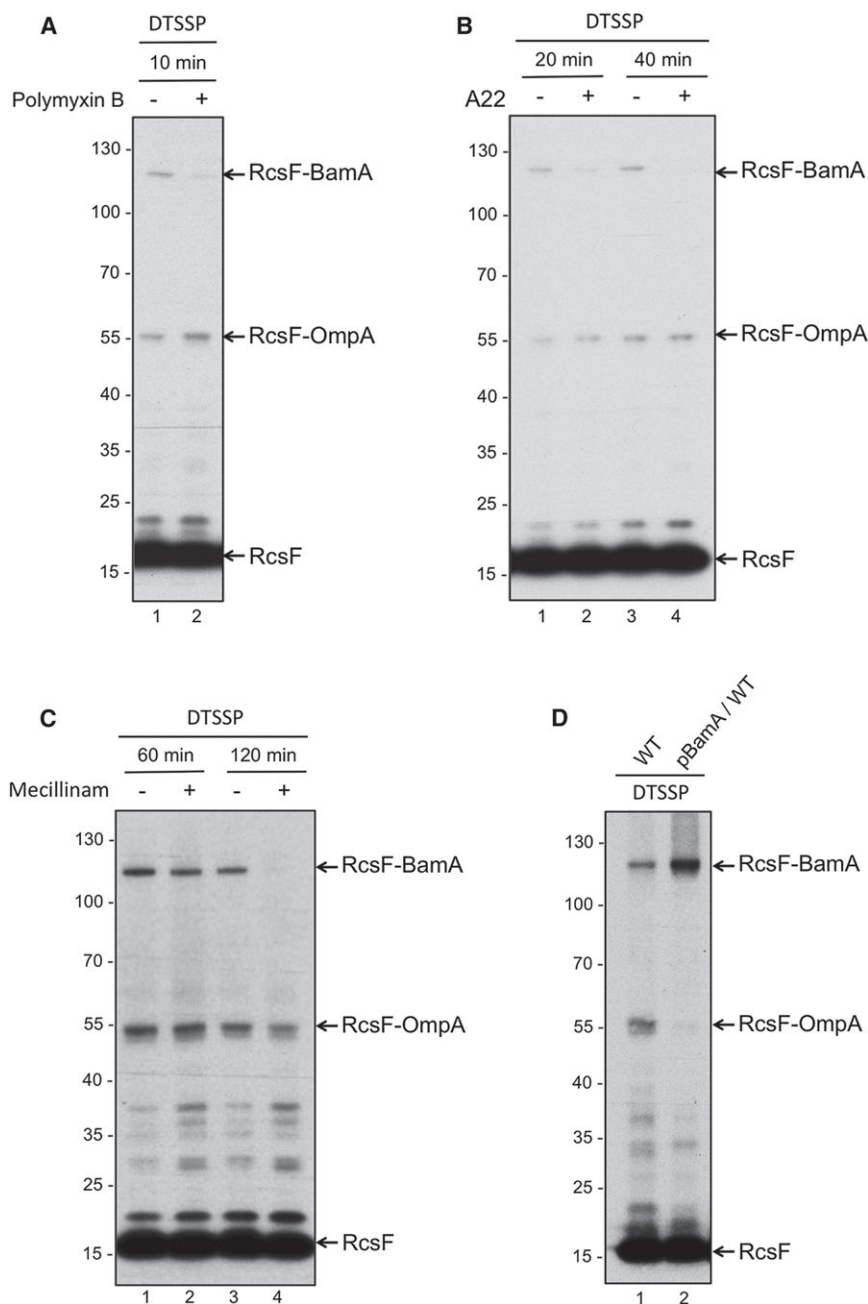


Figure 3. RcsF-BamA Is More Sensitive than RcsF-OmpA to Envelope Stress

(A–C) RcsF-BamA and RcsF-OmpA complex formation upon treatment with different cues sensed by RcsF (see also Figure S4A). (A) Cells were treated with 0.5 μ g/ml polymyxin B when they reached an OD₆₀₀ of 0.4, and samples were collected 10 min later, crosslinked with DTSSP and immunoblotted with an anti-RcsF antibody. (B and C) Cells were treated with mecillinam (0.3 μ g/ml) or A22 (5 μ g/ml) when they reached an OD₆₀₀ of 0.2, samples were collected at indicated time points after stress induction, and were subjected to DTSSP crosslinking and immunoblot. In all three stresses the RcsF-BamA complex disappeared when the Rcs system was activated (Figure S4A), whereas the RcsF-OmpA complex remained largely unaffected.

(D) BamA overexpression shifts all RcsF to BamA. In vivo DTSSP crosslinking of wild-type cells harboring an empty vector (pET3a) or a vector expressing BamA (pBamA). In all panels, DTSSP crosslinking and immunoblot were done as in Figure 2A, and a representative experiment is shown ($n = 3$ –4). See also Figure S4.

mutant that cannot produce UDP-D-glucose, a precursor for LPS and other surface-exposed sugars, and in which the Rcs system is constitutively turned on in an RcsF-dependent manner (Figure S4C) (Girgis et al., 2007). Similarly, the impact on the BamA-RcsF complex was stronger (Figure S4D). Thus, the BamA-RcsF complex was more responsive than the OmpA-RcsF complex, regardless of the stress applied. RcsF seems to be in a “locked” conformation with OmpA, which is not disrupted upon stress.

BamA is required for assembly of β -barrel proteins in the OM, including OmpA (Hagan et al., 2011). We postulated that BamA

assemble the OmpA-RcsF complex, with BamA funneling RcsF to OmpA.

Newly Synthesized RcsF Monitors the Activity of the Bam Machinery

We established that the BamA-RcsF interaction is key in the ability of RcsF to activate the Rcs system and in the assembly of the OmpA-RcsF complex. However, it remained unclear if the two events are connected, i.e., does formation of the OmpA-RcsF complex play a role in the ability of RcsF to sense stress? We reasoned that as only active BamA can form the OmpA-RcsF

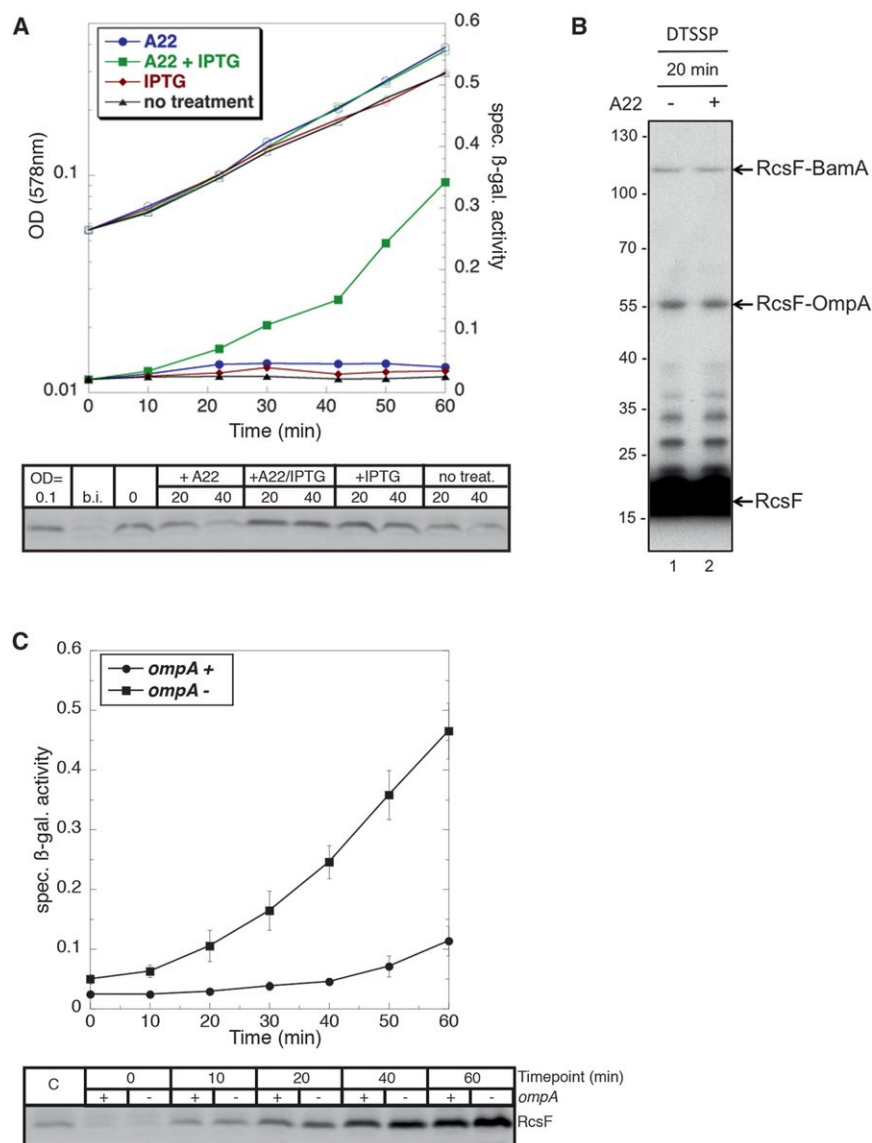


Figure 4. Newly Synthesized RcsF Senses BamA Activity

(A) Newly synthesized RcsF senses A22 stress. $\Delta rcsF$ cells carrying *rcsF* under an IPTG-inducible promoter on a low-copy vector (pNG162) and a higher-copy plasmid with *lacI^q* (pTrc-HIS2A) were grown without IPTG until an OD₅₇₈ of 0.05 (b.i. sample in western blot). 15 μ M IPTG was then added and cells were grown for 1 hr to reach RcsF steady-state levels (time point 0 in western blot) that were slightly higher than wild-type levels (RcsF being expressed from chromosome; first lane in western blot). At this point, cells were moved to media with/without inducer (IPTG) and/or the Rcs inducing cue (2 μ g/ml A22). Only cells with continued RcsF synthesis activated the Rcs system upon A22 stress. A chromosomal *rprA::lacZ* reporter was used to measure Rcs activity. Empty dots: OD; filled dots: specific β -gal activity. RcsF levels were monitored in parallel (bottom). A representative experiment of four replicates is shown.

(B) Preformed RcsF-BamA does not disassociate upon A22 stress. Cells were treated with chloramphenicol (300 μ g/ml) when they reached an OD₆₀₀ of 0.17, to block new protein synthesis, and subjected to A22 (5 μ g/ml) 10 min later (OD was then \sim 0.2). Samples were collected 20 min after stress induction and subjected to DTSSP cross-linking and immunoblot. A control strain was grown without drug for the same time. The levels of the BamA-RcsF complex remained constant.

(C) The Rcs system is more sensitive to RcsF levels in cells lacking OmpA. $\Delta rcsF$ and $\Delta rcsF \Delta ompA$ cells carrying *rcsF* under the same controllable expression system described in (A) were grown without inducer till an OD₅₇₈ of 0.3. At this point, IPTG (20 μ M) was added and Rcs activity (chromosomal *rprA::lacZ*) and RcsF levels (bottom) were monitored as a function of time. In $\Delta ompA$ cells the Rcs system was activated almost instantaneously, before RcsF reached the steady state levels observed in wild-type (with chromosomal *rcsF*). In contrast, in cells carrying OmpA, the Rcs system was activated only when RcsF reached a >3 -fold excess over the wild-type steady state levels, indicating that the presence of

OmpA increases the capacity of the cell to "store" RcsF. As BamA is required for funneling RcsF to OmpA, the cell senses this way the BamA activity—i.e., the ability to form the OmpA-RcsF complex. Error bars depict standard deviations ($n = 4$). See also Figure S5.

complex, if RcsF sensed the ability of BamA to funnel it to OmpA, then RcsF would actually monitor the activity of BamA. To test this hypothesis, we carefully dissected the interplay between RcsF, BamA, and OmpA. We first probed whether dissociation of preformed BamA-RcsF or inability of newly synthesized RcsF to bind to BamA upon stress triggers the Rcs system. To test this, we examined whether continuing RcsF synthesis was necessary for induction. We expressed RcsF to slightly higher steady-state levels than wild-type ($\sim 130\%$), using an IPTG inducible promoter. We then stressed cells with A22, and simultaneously either shut down RcsF expression or kept it at the same steady-state levels. Only cells with continued RcsF synthesis rapidly induced the Rcs system. Cells with only "old" RcsF, albeit to wild-type levels, could not induce the Rcs system

(Figure 4A). Similar results were obtained when we overexpressed RcsF and then completely shut down its expression, letting cells dilute RcsF to nearly wild-type levels before the A22 stress. In contrast to wild-type cells or to cells with reactivated RcsF expression, cells carrying only "old" RcsF (but at wild-type levels) could not activate the Rcs system (Figure S5A). Thus, new protein synthesis was required for RcsF to sense stress and activate the signaling cascade.

To obtain more direct evidence that induction of the Rcs system resulted from the inability of newly synthesized RcsF to bind to BamA upon stress, we induced the Rcs system with A22 shortly after stopping new protein synthesis and monitored the levels of the BamA-RcsF complex. We used A22 because it activates the Rcs system almost instantaneously and for a long

period of time (Figure S4A). In contrast, mecillinam-mediated activation was slow, while polymyxin B-mediated activation was short-lived (Figure S4A) (Farris et al., 2010). The BamA-RcsF complex levels remained unchanged when protein synthesis was stopped before A22 addition (Figure 4B), whereas they decreased when protein synthesis was ongoing (Figure 3B). This suggests that the newly arriving RcsF cannot bind to BamA when cells are stressed, leading to the activation of the system. Consistently the BamA-RcsF complex disappeared faster than OmpA-RcsF after adding A22 or polymyxin B under ongoing protein synthesis (Figures 3A–3B). This is because BamA could presumably keep funneling RcsF to OmpA to a certain degree, preventing OmpA-RcsF from disappearing with dilution-like kinetics. We also stopped new protein synthesis by chilling the cells, then added polymyxin B and probed RcsF-BamA and RcsF-OmpA complex formation after 10 min. Both complexes remained intact (Figure S5B), which is in agreement with the inability of preformed RcsF-BamA to respond to stress.

These results indicated that constant synthesis of RcsF is required for RcsF to act as a sensor, and supported a model in which activation results from newly synthesized RcsF being unable to bind BamA. However, it remained unclear whether irreversible sequestration of RcsF by BamA was sufficient to keep the Rcs system off or whether continuous funneling of RcsF to OmpA was also required. To discriminate between these two possibilities, we compared the levels of RcsF that were required to activate the system in $\Delta ompA$ and wild-type strains. While both strains have similar BamA levels (Figure S5C), BamA can funnel RcsF to OmpA only in the wild-type, thereby theoretically increasing its capacity for RcsF. We found that in the strain lacking *ompA*, the Rcs system was induced at a fraction (<80%) of wild-type RcsF levels, whereas wild-type cells could tolerate a ~3-fold increase in RcsF levels before inducing the Rcs system (Figure 4C). Thus, by funneling RcsF to OmpA, BamA increases its capacity for RcsF and maintains the Rcs system in an off state. This means that RcsF can monitor the capacity of BamA to assemble the OmpA-RcsF complex, which is presumably affected during stress. Since active BamA is required for OmpA-RcsF assembly (Figure 3D), RcsF senses this way the activity of the Bam machinery.

RcsF, a Sensitive but Robust-to-Noise Sensor of BamA Activity

Newly synthesized RcsF represents a small fraction of the RcsF pool, but is able to rapidly activate the system under stress. This led us to probe for the minimal RcsF protein levels required for activation. Taking advantage of the fact that periplasmic or IM-located RcsF constitutively activate the Rcs system (Farris et al., 2010), we established that RcsF_{peri} or RcsF_{IM} activated the Rcs system when they reached ~10% of the wild-type RcsF levels (Figures 5A and S6). Interestingly, protein abundance estimates based on ribosomal profiling indicate that RcsF (~3,100 copies/cell) is in ~10-fold excess over IgA (~220 copies/cell) and in similar amounts to BamA (~3,900 copies/cell) (Li et al., 2014). As RcsF activates the Rcs system via IgA, these numbers are consistent with the levels of RcsF required to remain exposed to the periplasm for activation.

Thus, RcsF is a sensitive sensor, able to trigger the Rcs response as soon as a small fraction escapes the BamA-OmpA pathway due to envelope perturbations.

This setup raised the possibility that small fluctuations of RcsF levels could trigger the Rcs phosphorelay, resulting into a leaky signaling system. This was not the case, as wild-type cells could tolerate significantly higher RcsF levels without triggering the Rcs response (Figure 4C) by funneling more RcsF to OmpA (Figure 5B). This suggests that during steady-state, BamA is not forming RcsF-OmpA at maximal capacity, thereby insulating the Rcs response from small fluctuations of RcsF levels. In comparison to RcsF and BamA, each cell contains ~210,000 copies of OmpA (Li et al., 2014), suggesting that OmpA levels are not the limiting factor in this process.

Portions of RcsF Are Displayed on the Cell Surface via OMPs

Although *E. coli* OM lipoproteins are considered to be all facing the periplasm, some were recently proposed to be surface-exposed (Zückert, 2014). Because the OmpA-RcsF complex was stable and unresponsive to stress, we hypothesized that the interaction between these two proteins may lead to partial exposure of RcsF on the surface. In this scenario, OmpA would be the vehicle for lipoprotein surface exposure, and BamA the means. To test this hypothesis, we performed immunofluorescence (IF) microscopy on intact and OM-permeabilized cells using antibodies specific for the signaling domain of RcsF. LamB_{3xFlag}, an abundant OM porin fused to a triple Flag tag at its periplasmic C terminus, was used as negative control (Figure S7A). RcsF was clearly labeled in intact cells, while LamB_{3xFlag} was only marginally labeled in the same cells (Figure 6A). Similar results were obtained with immunodot blotting on intact cells without fixation (Figure S7B). Therefore, we propose that the signaling domain of RcsF is at least partially exposed on the cell surface.

We next tested whether surface exposure of RcsF required OmpA. We were unable to address this question using a $\Delta ompA$ mutant, as deleting *ompA* rendered the OM permeable to antibodies, making it impossible to obtain reliable results. Instead, we used BamA overexpressing cells, in which the OmpA-RcsF complex was almost absent (Figure 3D). In this case, the surface-exposed RcsF decreased significantly (Figure 6B) without an overall change in RcsF levels (Figure S7C), suggesting that RcsF reaches the surface at least partially via OmpA. Conversely, overexpression of RcsF that increased OmpA-RcsF levels (Figure 5B) resulted in more RcsF being detected on the cell surface (Figure 6C). Altogether these results suggest that RcsF reaches the cell surface mainly via OmpA, but possibly also through other OMPs (see Discussion).

DISCUSSION

RcsF Senses the Bam Machinery Activity

After IM translocation, β -barrels are ushered by chaperones through the periplasm to the Bam machinery, which folds and inserts them into the OM (Goemans et al., 2014). To monitor malfunctioning at different levels of this multistep process, the cell would need multiple signal transduction systems. We already

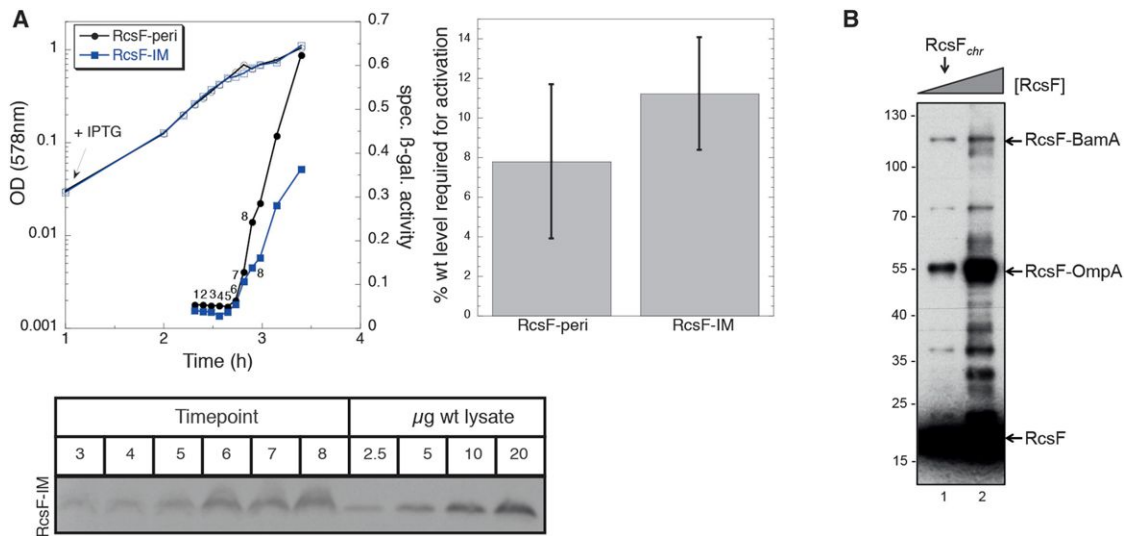


Figure 5. RcsF Is a Sensitive but Robust-to-Noise Sensor of BamA Activity

(A) Only a fraction of RcsF is required to be in the periplasm for Rcs activation. $\Delta rcsF$ cells carrying RcsF_{IM} or RcsF_{peri} under an IPTG-inducible promoter on a low-copy vector (pSC202) and a higher-copy plasmid encoding *lacI^q* (pREP4) were grown for three generations in LB and before adding inducer (100 μ M IPTG). RcsF protein levels (bottom) and Rcs activity (top left; chromosomal *rprA::lacZ* fusion) were closely monitored onward. Note that in this setup, no matter how much IPTG was added, or when added, *rcsF* expression remained undetectable until cells reached an OD₅₇₈ of ~ 0.6 . Quantification of RcsF_{IM} or RcsF_{peri} protein levels at the time point of Rcs activation is shown at the top right. Error bars depict standard deviation ($n = 3$). The time point of activation was considered as the point at which a linear curve fitted on specific β -gal activity versus time crossed the basal activity, minus 3 min required for β -gal synthesis and folding. For quantifying RcsF levels, we always ensured that the signal detected from cells expressing RcsF_{IM} or RcsF_{peri} (40 μ g) was within linear range by loading a titration of total protein extracts from wild-type cells (2.5–20 μ g). An example western blot is shown (bottom). Empty dots: OD; filled dots: specific β -gal activity. The full gel can be seen in Figure S6.

(B) The capacity of BamA to form the OmpA-RcsF complex is not maxed out in wild-type cells. Increasing RcsF expression resulted into more OmpA-RcsF complex being formed, but the levels of the BamA-RcsF complex remained largely unchanged. Thus, in nonstressed cells, BamA has the ability to funnel more RcsF to OmpA. In lane 1 the wild-type levels of the OmpA-RcsF and BamA-RcsF complexes are shown. In lane 2, RcsF was expressed in $\Delta rcsF$ cells carrying *rcsF* under an IPTG-inducible promoter on a low-copy vector, pSC202. In the absence of *lacI^q*, the RcsF steady-state levels were ~ 3 - to 4-fold higher than in the wild-type. DTSSP crosslinking and immunoblot were performed as described in Figure 2A, and a representative experiment is shown ($n = 3$).

know that accumulation of unassembled OMPs in the periplasm is the primary signal for the σ^E stress response (Walsh et al., 2003). We now report that the activity of the Bam machinery is monitored by the Rcs system through RcsF. BamA interacts with RcsF and, when active, funnels it to OmpA. When bound to BamA or OmpA, RcsF is occluded from IgA and cannot activate the Rcs system. Yet, BamA cannot sequester all RcsF molecules and funneling of newly synthesized RcsF to OmpA is necessary for maintaining the Rcs system off. This is especially important because preformed BamA-RcsF does not disassociate upon stress, and only newly arriving RcsF can sense stress. Thus, this constant flow of RcsF from BamA to OmpA is what defines the availability of BamA and what RcsF is sensing. Stress conditions impair BamA availability for newly arriving RcsF, which ends up facing the periplasm, free to activate the Rcs cascade (Figure 7).

In addition to its well-known activity in OMPs assembly, we report that BamA funnels RcsF to OmpA and other OMPs. Since functional Bam machinery is required for both events, we suggest that they are coupled (Figure 7), which implies that RcsF senses by default both activities. Further structure-function analysis will be required for deciphering if and how the two events are connected and how RcsF intervenes.

An interesting feature of the Rcs system is that RcsF is in ~ 10 -fold excess over its downstream partner IgA (Li et al., 2014), despite the two forming a 1:1 complex (Figure 1C). This results into only a fraction of RcsF being required for fully activating the Rcs system (Figure 5A). The cell presumably maintains RcsF in excess over IgA to efficiently monitor the Bam machinery, which is present at a similar copy/cell ratio as RcsF (Li et al., 2014). At the same time, the steady-state RcsF levels are kept low enough to prevent activation of the Rcs system by small fluctuations. Indeed, a ~ 3 -fold increase in RcsF levels was required for the Rcs system to be activated without stress (Figure 4C). An interesting hypothesis that we are currently pursuing is that RcsF levels are optimized for high sensitivity and low noise.

How can PG and OM stress affect Bam activity? Although PG perturbations could affect the journey of RcsF bound to the lipoprotein-specific chaperone LolA through the porous PG layer, we did not see RcsF accumulating in the periplasm upon mecillinam or A22 treatment. On the other hand, transport of the bulkier BamA may be more impaired, creating a bottleneck in BamA availability/activity. Alternatively the POTRA domains of BamA that extend deep into the periplasm could be affected by changes in PG integrity, with direct consequences

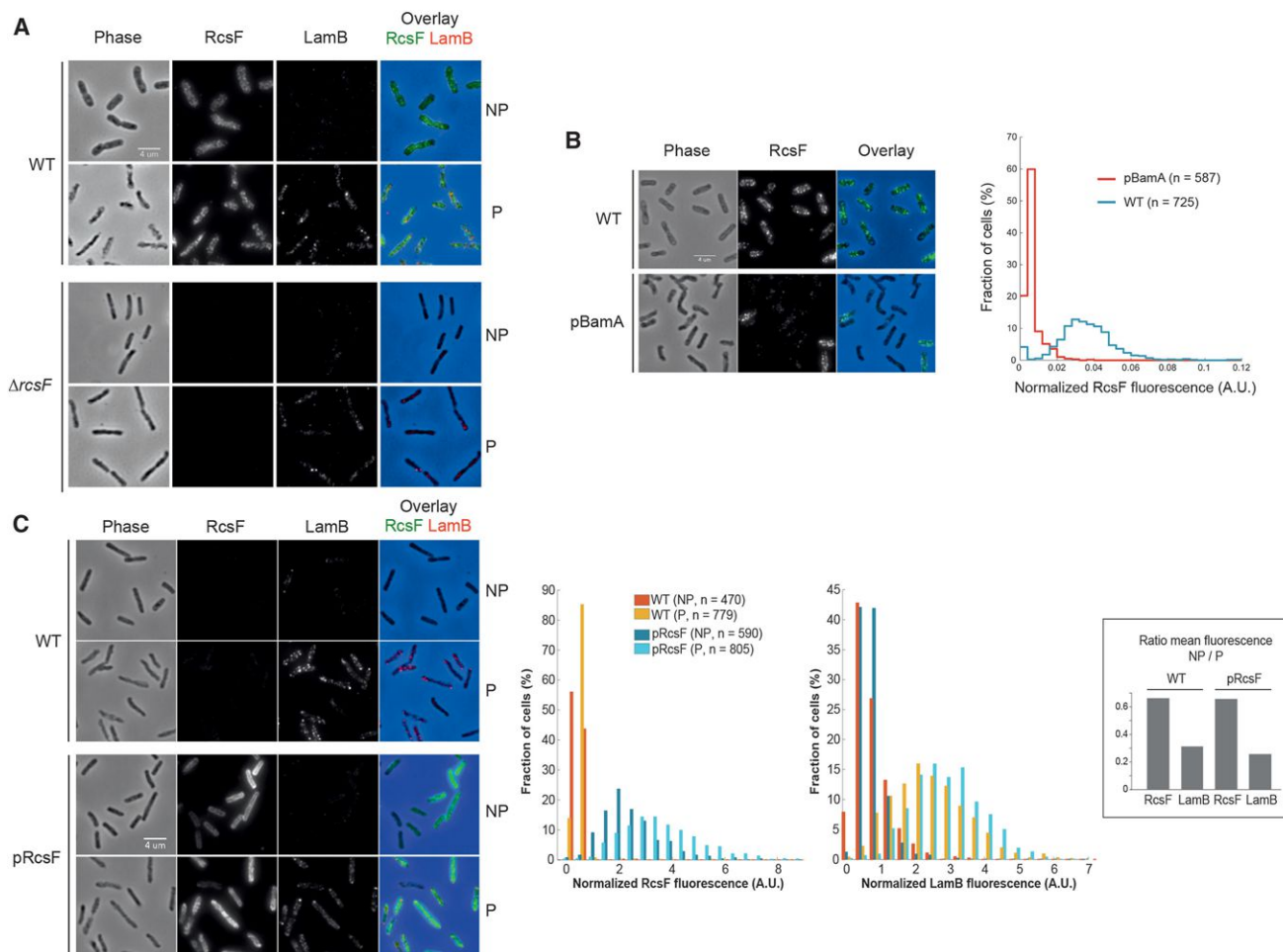


Figure 6. Portions of RcsF Are Surface-Exposed

(A) Wild-type and Δ rcsF cells were probed for RcsF and LamB_{3xFlag} localization by IF microscopy using anti-RcsF/-Flag antibodies, with/without cell permeabilization. LamB_{3xFlag}, expressed from a plasmid, serves as a permeabilization marker—the 3xFlag tag is fused to the periplasmic C terminus (Figure S7A). The Δ rcsF strain is used as specificity control for the anti-RcsF antibody. Phase contrast, fluorescence signals and overlay images (green: RcsF, red: LamB_{3xFlag}) are shown for representative cells. Scale bar, 4 μ m. P, permeabilized; NP, nonpermeabilized. Unlike LamB, RcsF is detected even on nonpermeabilized wild-type cells.

(B) BamA overexpression reduces RcsF surface exposure. RcsF was visualized by IF as described above in nonpermeabilized cells carrying an empty vector (WT) or a vector overexpressing BamA (pBamA). Phase contrast, fluorescence signal and overlay images are shown for representative cells. Scale bar: 4 μ m. Right: a distribution of the total fluorescence intensity per cell, normalized by cell area, is shown for populations of WT (blue) and pBamA (red) cells. AU, arbitrary units; n, number of cells. Significantly less RcsF is detected on the surface when BamA is overexpressed.

(C) RcsF overexpression increases its surface exposure. RcsF and LamB_{3xFlag} were visualized by IF as described above in wild-type cells containing an empty vector, pBAD33 (WT), and Δ rcsF cells carrying pSC216 (pBAD33-RcsF). Phase contrast, fluorescence signal and overlay images are shown for representative cells. Scale bar, 4 μ m. Note that the RcsF-associated signal is not visible in nonpermeabilized WT cells because of scaling applied—to avoid saturation of the RcsF signal in pRcsF cells. Middle, distributions of the total fluorescence intensity (associated with RcsF, left, or LamB_{3xFlag}, right) per cell, normalized by cell area, are shown for populations of wild-type and pRcsF cells imaged on the left. Inset: the ratio of the mean value of normalized fluorescence associated with RcsF or LamB_{3xFlag} in nonpermeabilized and permeabilized (NP/P) wild-type and pRcsF cells is depicted. Abbreviations are like above. A significantly higher fraction of RcsF could be labeled from outside, compared to LamB. Although more RcsF was detected on the surface of cells overexpressing RcsF than in WT cells, the fraction of surface-exposed RcsF (NP/P ratio) remained similarly high. See also Figure S7.

on the Bam activity. Defects in LPS composition and assembly could also affect the RcsF journey in many ways, as they vastly reorganize the OM and periplasm (Sperandeo et al., 2008). Further work is required to mechanistically dissect how particular envelope stresses impair the availability and activity of BamA.

The Bam Machinery Exports RcsF to the Cell Surface

Although the general view has been that *E. coli* OM lipoproteins face the periplasm (Okuda and Tokuda, 2011), a handful were recently reported to be surface exposed (Zückert, 2014). Yet no machinery has been identified that would allow such translocation through the OM bilayer. Here, we report that BamA allows

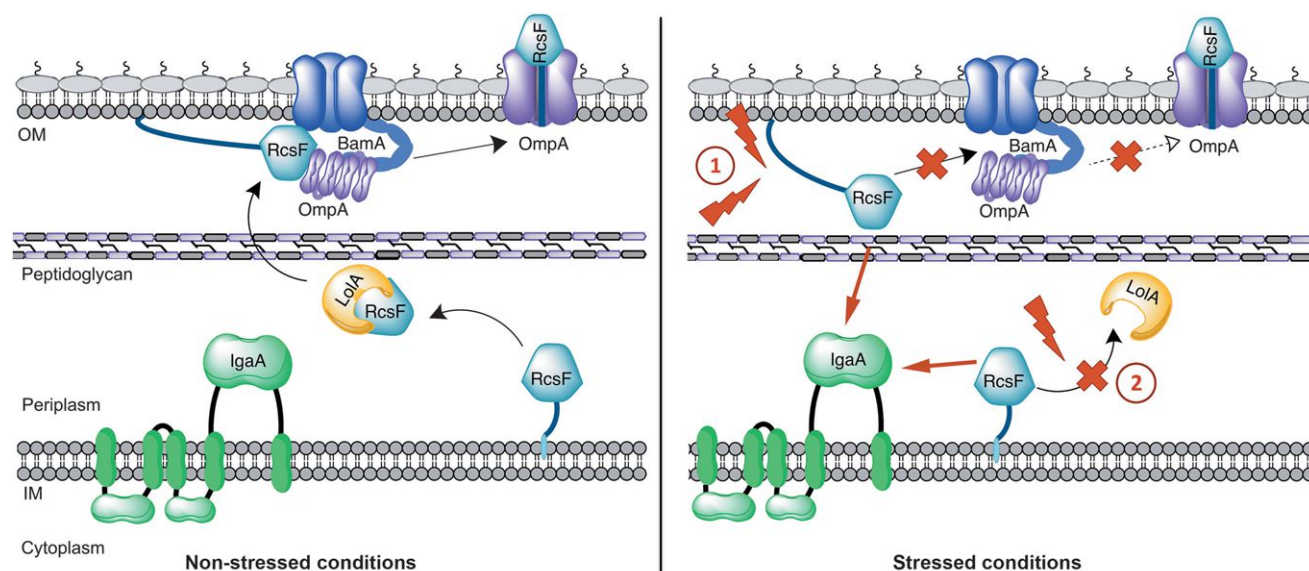


Figure 7. RcsF Monitors the Journey of Lipoproteins from the IM to the Cell Surface

A cartoon depicting our model: RcsF acts as a sensor of lipoprotein transport to the OM by the Lol system, and export to the cell surface by the Bam machinery. RcsF, as other OM lipoproteins, is transported to the inner leaflet of the OM by the chaperone LolA. In the absence of stress (left), newly transported RcsF interacts with BamA, the key component of the β -barrel assembly machinery. BamA assembles a complex between RcsF and OmpA, an abundant β -barrel protein, in way that RcsF is displayed on the cell surface. Once engaged in interactions with BamA and/or OmpA, RcsF is occluded from IgaA, the downstream component of the Rcs signaling cascade located in the IM. Upon OM or PG-related stress (right), newly transported RcsF fails to bind BamA, possibly because the activity of BamA is perturbed, creating a bottleneck in the availability of “cargo-free” BamA (1). This results into newly synthesized RcsF being exposed to the periplasm, where it binds to IgaA and triggers the Rcs system. RcsF also gains access to IgaA when lipoproteins accumulate in the IM due to failures in lipoprotein transport to the OM (2). Note that RcsF still reaches the OM under OM or PG-related stresses (Figure S4B), so IM accumulation is not the reason for Rcs activation in these conditions. Although RcsF and BamA are synthesized at the same rates (Li et al., 2014), only a small fraction of unbound RcsF in the periplasm is enough for activating the Rcs system. Thus, we propose that RcsF is a sensitive sensor that monitors the availability, and thereby the ability, of the Bam machinery to assemble OMPs and/or target lipoproteins to the outer surface via OMPs.

RcsF to reach the surface by funneling it to the β -barrel OmpA. It is plausible that the BamA-mediated formation of lipoprotein- β -barrel complexes is a more general mechanism of lipoprotein export to the surface, but further experimentation is needed to establish this. As the Bam machine is highly conserved among Gram-negative bacteria, this would explain why surface-exposed OM lipoproteins are exported to the outside when expressed in heterologous systems (Arnold et al., 2014; Pride et al., 2013).

OmpA is the major β -barrel acting as terminal acceptor of RcsF. Not only most RcsF binds OmpA (Figures 2B and S3E and S3F), but also the steady-state Rcs activity is higher in $\Delta ompA$ cells than in cells deleted for other β -barrels (Figure 2D). Moreover, BamA has lower capacity for RcsF in $\Delta ompA$ cells (Figure 4C). Yet our MS and Rcs-activity data (Table S2A and Figure 2D) indicated that RcsF is also funneled to other abundant β -barrels, such as OmpC and OmpF (~165,000 and 90,000 copies/cell, respectively [Li et al., 2014]). The cellular levels of the 3 OMPs cannot explain alone the preference of BamA for funneling RcsF to OmpA, suggesting that a more selective process, which remains to be discovered, is at play. The redundancy of OMPs as terminal acceptors of RcsF could explain why $\Delta ompA$ cells or cells missing 2 OMPs have the Rcs system still only partially activated, in comparison to cells where BamA is depleted. Interestingly, if many OMPs are used as terminal RcsF acceptors, this could

mean that BamA, on its own, has very limited capacity to sequester RcsF, and has to always funnel new RcsF to OMPs for keeping the Rcs system off. As BamA levels are slightly higher than that of RcsF (Li et al., 2014), there may be more (lipoprotein) substrates competing with RcsF for BamA.

We have shown that at least portions of the RcsF signaling domain reach the cell surface with formation of the OmpA-RcsF complex being required for this (Figure 6). Indeed, in the presence of excess nonfunctional BamA, the OmpA-RcsF complex became undetectable (Figure 3D) and the surface-exposed RcsF significantly decreased. Although these results indicate that RcsF can use OmpA to reach the surface, they do not exclude that other OMPs are also used, as overexpression of nonfunctional BamA will also prevent the formation of any OMP-RcsF complex.

RcsF interacts with OmpA via its N-terminal linker and the tip of the signaling domain (Figures 2C and S3). A simple model to explain our results is that the N-terminal linker of RcsF traverses the OmpA pore to allow (portions of) the globular domain to locate outside the OM (Figure 7). This would not be unprecedented as OM lipoproteins with their entire globular domain present on the surface have been reported, such as the *Vibrio cholerae* lysophospholipase VolA (Pride et al., 2013). As OmpA-RcsF is a dead-end complex for the signaling role of RcsF, the physiological role of RcsF when bound to OmpA is

enigmatic. Additional work will be required to clarify how OmpA and RcsF interact and the role of RcsF in this complex.

How can RcsF use a β -barrel such as OmpA to access the surface? OM porins act as gates for peptides coming from outside (Housden et al., 2013) and for periplasmic proteins secreted by the cell, such as YebF (Prehna et al., 2012). The lipoprotein LptE was also recently shown to reside inside the β -barrel LptD, presumably acting as a controllable plug for the LPS assembly machinery (Freinkman et al., 2011). Thus, it is not uncommon that a β -barrel pore can accommodate a polypeptide. OmpA, in one of its two known conformations, forms a 16-stranded β -barrel structure with a large pore (Reusch, 2012). This conformation could accommodate a disordered segment such as the RcsF linker. In its second conformation, which has been proposed to be an intermediate state, OmpA assumes a 2-domain structure, with a smaller N-terminal β -barrel and a C-terminal periplasmic domain interacting with the PG (Reusch, 2012). In this conformation, the β -barrel diameter is too small for a polypeptide, but an OmpA-RcsF interaction at this stage could be an important intermediate for the funneling of RcsF from BamA.

Finally, we detected an RcsF-Lpp complex (Figure S2D). As this interaction was not recapitulated with any of the 25 pBpa-containing RcsF variants, we deduced that it might be indirect. This would be consistent with the very high abundance of Lpp and its shared localization with RcsF at both OM leaflets (Cowles et al., 2011). In addition, the absence of Lpp did not affect the RcsF-BamA and RcsF-OmpA interactions (Figure S2D). It remains to be tested if Lpp has any direct effect on Rcs signaling.

Integrating Envelope Stresses: RcsF Monitors the Journey of Lipoproteins through the Envelope

There are ~100 lipoproteins in *E. coli*. The vast majority is localized in the OM. Although the function of most is unknown, some are components of essential OM assembly machineries (Silhavy et al., 2010) and others regulate core envelope processes (Paradis-Bleau et al., 2010; Typas et al., 2010; Uehara et al., 2010). Thus lipoprotein targeting is vital for the cell.

OM lipoproteins are escorted across the periplasm by the essential chaperone LolA (Okuda and Tokuda, 2011). RcsF senses defects in: (1) phosphatidylglycerol biosynthesis (Shiba et al., 2004), which is required for lipoprotein maturation; and (2) in the LolA-mediated transport of lipoproteins to the OM, presumably because it gets stuck in the IM when LolA's function is impaired (Tao et al., 2012), gaining access to IgaA. Rcs activation resulting from RcsF accumulation in the IM leads to higher *lolA* expression, creating a feedback loop to fix the damage (Tao et al., 2012).

For RcsF, and at least a few other lipoproteins, the journey does not end at the inner leaflet of the OM, as they are finally translocated to the cell surface. As we have shown here, it is the Bam machine that mediates the export of RcsF to the surface by inserting it into β -barrels such as OmpA. Malfunctioning of this process results into newly translocated RcsF remaining exposed in the periplasm, where it can reach IgaA and trigger the signaling cascade (Figure 7). Therefore, RcsF also monitors the ability of BamA to insert OM lipoproteins to OMPs. Altogether this means that Rcs can sense the entire lipoprotein journey across the en-

velope, from maturation to OM exposure, adjusting the envelope composition in response to failures at any step.

Rcs, a Complex Signal Transduction System

Rcs is one of the most complex signaling systems known in bacteria with key steps remaining unresolved. We have shown that RcsF interacts with the large periplasmic domain of IgaA, which likely triggers the signaling cascade. As the two membranes are separated by ~200 Å, it remains to be determined how this interaction occurs. RcsF has an intrinsically disordered 31 amino acid-long N-terminal linker. It is likely that, when extended, this region allows RcsF to reach the large periplasmic domain of IgaA. The OM lipoprotein LpoB uses a similar configuration to access its IM counterpart, PBP1B (Egan et al., 2014).

It also remains to be proven whether the RcsF-IgaA interaction is sufficient for conveying the signal downstream and activating the Rcs cascade; our genetic data that put IgaA downstream of RcsF strongly suggest so. How IgaA itself mechanistically controls the Rcs phosphorelay, whether it directly interacts with the other IM components RcsC and RcsD, and whether it plays additional roles in the cell remain unknown and will all be fields of future research. Moreover, further work will be required to elucidate how the few genetic perturbations that activate the Rcs system independently of RcsF (Majdalani and Gottesman, 2005, 2007) are sensed by the system.

CONCLUSIONS

We elucidated how the OM lipoprotein RcsF senses stress and talks to the downstream signaling cascade. RcsF monitors the activity of the machinery for OM β -barrel assembly, Bam, triggering the signaling cascade when Bam is malfunctioning. Moreover, we identified the formation of complexes between RcsF and the β -barrel OmpA as a novel mechanism for lipoprotein translocation through the bacterial OM. We propose that this may be a conserved system for lipoprotein export. Although many of the molecular details of both processes described here remain to be fully elucidated, these findings generate a number of intriguing hypotheses on the mechanisms that the cell uses to sense the activity of the protein machineries that build its envelope.

EXPERIMENTAL PROCEDURES

Bacterial Strains, Media, and Plasmids

Cells were grown in LB at 37°C and, when necessary, growth media were supplemented with spectinomycin (50–100 µg/ml), ampicillin (100–200 µg/ml), chloramphenicol (20–25 µg/ml), or kanamycin (50 µg/ml). The bacterial strains and plasmids used in this study are listed in Tables S3 and S4, respectively, and information on their construction is provided in Extended Experimental Procedures.

In Vitro RcsF-IgaA Binding

RcsF with a C-terminal 6-Histidine tag (RcsF-His) and an untagged version of the periplasmic IgaA domain were purified as described in Extended Experimental Procedures. RcsF-His (0.15 nmol) was coupled to 20 µl Talon beads and washed with PD buffer (25 mM Tris [pH 7.5], 200 mM NaCl, 10% glycerol) to remove residual RcsF-His. IgaA was then added to the RcsF (2.5 µM) containing Talon beads in a concentration range: 0.375–10 µM (assay volume = 60 µl; 0.625–10 µM range is shown in Figure 1C). The RcsF-IgaA suspension

was incubated for 15 min at room temperature and pelleted by brief centrifugation. Half of the supernatant was aspirated to quantify unbound IgA by SDS-PAGE. The pellet was washed with 500 μ l PD buffer and half was also analyzed by SDS-PAGE to quantify the pulled-down fraction of IgA.

In Vivo DTSSP Crosslinking

In vivo chemical crosslinking experiments were performed as described by Thanabalu et al. (1998) with some modifications. The detailed procedures are described in Extended Experimental Procedures.

In Vivo Site-Specific Photocrosslinking

Site-specific photocrosslinking was performed essentially as described by Okuda et al. (2012) with some modifications. The detailed procedures are described in Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.045>.

AUTHOR CONTRIBUTIONS

S.H.C., J.S., C.P., A.T., and J.F.C. conceived this study; S.H.C., J.S., C.P., M.Z., M.B., A.T., and J.F.C. designed research; S.H.C., J.S., C.P., M.Z., P.R., A.A., G.L., M.B., A.-K.H., P.L., C.V.d.H., D.V., and A.T. performed all experiments (S.H.C., J.S. performed all DTSSP experiments, except for the pull down-MS for IgA-peri which were performed by C.P. and M.Z.; S.H.C. performed all photocrosslinking experiments; C.P., M.Z., and A.K. performed all β -gal assays; C.P. performed the quantitative western blotting; M.B. and A.K. performed the in vitro binding assays; P.R., A.A., and G.L. performed the IF microscopy; J.S. performed the dot blotting; A.T. performed the *igaA* suppressor screen; S.H.C., J.S., C.P., M.Z., P.R., A.A., G.L., M.B., D.V., A.T., and J.F.C. interpreted the data; A.T. and J.F.C. wrote the manuscript; S.H.C., J.S., C.P., A.A., G.L. prepared the figures; S.H.C., J.S. and C.P. edited the manuscript; A.T. and J.F.C. supervised all aspects of the project.

ACKNOWLEDGMENTS

We are grateful to C.A. Gross, K.C. Huang, and N. Ruiz for critically reading the manuscript and providing feedback. We thank H. Mori (Nara, Japan), S. Gottesman (NIH, USA), D. Kahne (Harvard, USA), and T. Silhavy (Princeton, USA) for providing strains. We thank the members of the two labs for helpful discussions, A. Boujdat, G. Herinckx, A. Koumoutsis, and M. Pruteanu for assistance in experiments and D. Colau for antibody purification. M.S. was performed at the Proteomics Core Facilities of EMBL and of the de Duve Institute. J.S. is a FRIA research fellow, C.P. is an EMBO long-term fellow, P.L. is a "Chargée de Recherche" and J.F.C. is "Maître de Recherche" of the FRS-FNRS. This work was supported by grants from the FRS-FNRS and the Interuniversity Attraction Pole Program-Belgian Science Policy (network P7/44) to JFC, and the Sofia Kovalevskaja Award of the Alexander von Humboldt Foundation and EMBL internal funding to A.T.

Received: July 11, 2014

Revised: October 6, 2014

Accepted: November 24, 2014

Published: December 18, 2014

REFERENCES

Aoki, S.K., Malinverni, J.C., Jacoby, K., Thomas, B., Pamma, R., Trinh, B.N., Remers, S., Webb, J., Braaten, B.A., Silhavy, T.J., and Low, D.A. (2008). Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. *Mol. Microbiol.* 70, 323–340.

Arnold, M.F., Caro-Hernandez, P., Tan, K., Runti, G., Wehmeier, S., Scocchi, M., Doerrler, W.T., Walker, G.C., and Ferguson, G.P. (2014). Enteric YaiW is

a surface-exposed outer membrane lipoprotein that affects sensitivity to an antimicrobial peptide. *J. Bacteriol.* 196, 436–444.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 0008.

Cano, D.A., Domínguez-Bernal, G., Tierrez, A., García-Del Portillo, F., and Casadesús, J. (2002). Regulation of capsule synthesis and cell motility in *Salmonella enterica* by the essential gene *igaA*. *Genetics* 162, 1513–1523.

Castanié-Cornet, M.P., Cam, K., and Jacq, A. (2006). RcsF is an outer membrane lipoprotein involved in the RcsCDB phosphorelay signaling pathway in *Escherichia coli*. *J. Bacteriol.* 188, 4264–4270.

Cowles, C.E., Li, Y., Semmelhack, M.F., Cristea, I.M., and Silhavy, T.J. (2011). The free and bound forms of Lpp occupy distinct subcellular locations in *Escherichia coli*. *Mol. Microbiol.* 79, 1168–1181.

Domínguez-Bernal, G., Pucciarelli, M.G., Ramos-Morales, F., García-Quintanilla, M., Cano, D.A., Casadesús, J., and García-del Portillo, F. (2004). Repression of the RcsC-YojN-RcsB phosphorelay by the IgA protein is a requisite for *Salmonella* virulence. *Mol. Microbiol.* 53, 1437–1449.

Egan, A.J., Jean, N.L., Koumoutsis, A., Bougault, C.M., Biboy, J., Sassine, J., Solovyova, A.S., Breukink, E., Typas, A., Vollmer, W., and Simorre, J.P. (2014). Outer-membrane lipoprotein LpoB spans the periplasm to stimulate the peptidoglycan synthase PBP1B. *Proc. Natl. Acad. Sci. USA* 111, 8197–8202.

Evans, K.L., Kannan, S., Li, G., de Pedro, M.A., and Young, K.D. (2013). Eliminating a set of four penicillin binding proteins triggers the Rcs phosphorelay and Cpx stress responses in *Escherichia coli*. *J. Bacteriol.* 195, 4415–4424.

Farris, C., Sanowar, S., Bader, M.W., Pfuetzner, R., and Miller, S.I. (2010). Antimicrobial peptides activate the Rcs regulon through the outer membrane lipoprotein RcsF. *J. Bacteriol.* 192, 4894–4903.

Freinkman, E., Chng, S.S., and Kahne, D. (2011). The complex that inserts lipopolysaccharide into the bacterial outer membrane forms a two-protein plug-and-barrel. *Proc. Natl. Acad. Sci. USA* 108, 2486–2491.

Girgis, H.S., Liu, Y., Ryu, W.S., and Tavazoie, S. (2007). A comprehensive genetic characterization of bacterial motility. *PLoS Genet.* 3, 1644–1660.

Goemans, C., Denoncin, K., and Collet, J.F. (2014). Folding mechanisms of periplasmic proteins. *Biochim. Biophys. Acta* 1843, 1517–1528.

Gogol, E.B., Rhodius, V.A., Papenfort, K., Vogel, J., and Gross, C.A. (2011). Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proc. Natl. Acad. Sci. USA* 108, 12875–12880.

Guo, M.S., Updegrove, T.B., Gogol, E.B., Shabalina, S.A., Gross, C.A., and Storz, G. (2014). MicL, a new σ E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev.* 28, 1620–1634.

Hagan, C.L., Kim, S., and Kahne, D. (2010). Reconstitution of outer membrane protein assembly from purified components. *Science* 328, 890–892.

Hagan, C.L., Silhavy, T.J., and Kahne, D. (2011). β -Barrel membrane protein assembly by the Bam complex. *Annu. Rev. Biochem.* 80, 189–210.

Housden, N.G., Hopper, J.T., Lukyanova, N., Rodríguez-Larrea, D., Wojdyla, J.A., Klein, A., Kaminska, R., Bayley, H., Saibil, H.R., Robinson, C.V., and Kleanthous, C. (2013). Intrinsically disordered protein threads through the bacterial outer-membrane porin OmpF. *Science* 340, 1570–1574.

Laubacher, M.E., and Ades, S.E. (2008). The Rcs phosphorelay is a cell envelope stress response activated by peptidoglycan stress and contributes to intrinsic antibiotic resistance. *J. Bacteriol.* 190, 2065–2074.

Leverrier, P., Declercq, J.P., Denoncin, K., Vertommen, D., Hiniker, A., Cho, S.H., and Collet, J.F. (2011). Crystal structure of the outer membrane protein RcsF, a new substrate for the periplasmic protein-disulfide isomerase DsbC. *J. Biol. Chem.* 286, 16734–16742.

Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635.

- Lima, S., Guo, M.S., Chaba, R., Gross, C.A., and Sauer, R.T. (2013). Dual molecular signals mediate the bacterial response to outer-membrane stress. *Science* **340**, 837–841.
- Majdalani, N., and Gottesman, S. (2005). The Rcs phosphorelay: a complex signal transduction system. *Annu. Rev. Microbiol.* **59**, 379–405.
- Majdalani, N., and Gottesman, S. (2007). Genetic dissection of signaling through the Rcs phosphorelay. *Methods Enzymol.* **423**, 349–362.
- Okuda, S., and Tokuda, H. (2011). Lipoprotein sorting in bacteria. *Annu. Rev. Microbiol.* **65**, 239–259.
- Paradis-Bleau, C., Markovski, M., Uehara, T., Lupoli, T.J., Walker, S., Kahne, D.E., and Bernhardt, T.G. (2010). Lipoprotein cofactors located in the outer membrane activate bacterial cell wall polymerases. *Cell* **143**, 1110–1120.
- Prehna, G., Zhang, G., Gong, X., Duszyk, M., Okon, M., McIntosh, L.P., Weiner, J.H., and Strynadka, N.C. (2012). A protein export pathway involving *Escherichia coli* porins. *Structure* **20**, 1154–1166.
- Pride, A.C., Herrera, C.M., Guan, Z., Giles, D.K., and Trent, M.S. (2013). The outer surface lipoprotein VolA mediates utilization of exogenous lipids by *Vibrio cholerae*. *MBio* **4**, e00305–e00313.
- Reusch, R.N. (2012). Biogenesis and functions of model integral outer membrane proteins: *Escherichia coli* OmpA and *Pseudomonas aeruginosa* OprF. *FEBS J.* **279**, 893.
- Rhodium, V.A., Suh, W.C., Nonaka, G., West, J., and Gross, C.A. (2006). Conserved and variable functions of the sigmaE stress response in related genomes. *PLoS Biol.* **4**, e2.
- Shiba, Y., Yokoyama, Y., Aono, Y., Kiuchi, T., Kusaka, J., Matsumoto, K., and Hara, H. (2004). Activation of the Rcs signal transduction system is responsible for the thermosensitive growth defect of an *Escherichia coli* mutant lacking phosphatidylglycerol and cardiolipin. *J. Bacteriol.* **186**, 6526–6535.
- Silhavy, T.J., Kahne, D., and Walker, S. (2010). The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2**, a000414.
- Sperandeo, P., Lau, F.K., Carpentieri, A., De Castro, C., Molinaro, A., Dehò, G., Silhavy, T.J., and Polissi, A. (2008). Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of *Escherichia coli*. *J. Bacteriol.* **190**, 4460–4469.
- Tao, K., Narita, S., and Tokuda, H. (2012). Defective lipoprotein sorting induces *lta* expression through the Rcs stress response phosphorelay system. *J. Bacteriol.* **194**, 3643–3650.
- Thanabalu, T., Koronakis, E., Hughes, C., and Koronakis, V. (1998). Substrate-induced assembly of a contiguous channel for protein export from *E. coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *EMBO J.* **17**, 6487–6496.
- Typas, A., Nichols, R.J., Siegele, D.A., Shales, M., Collins, S.R., Lim, B., Bräber, H., Yamamoto, N., Takeuchi, R., Wanner, B.L., et al. (2008). High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat. Methods* **5**, 781–787.
- Typas, A., Banzhaf, M., van den Berg van Saparoea, B., Verheul, J., Biboy, J., Nichols, R.J., Zietek, M., Beilharz, K., Kannenberg, K., von Rechenberg, M., et al. (2010). Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell* **143**, 1097–1109.
- Typas, A., Banzhaf, M., Gross, C.A., and Vollmer, W. (2012). From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat. Rev. Microbiol.* **10**, 123–136.
- Uehara, T., Parzych, K.R., Dinh, T., and Bernhardt, T.G. (2010). Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. *EMBO J.* **29**, 1412–1422.
- Walsh, N.P., Alba, B.M., Bose, B., Gross, C.A., and Sauer, R.T. (2003). OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell* **113**, 61–71.
- Zhang, M., Lin, S., Song, X., Liu, J., Fu, Y., Ge, X., Fu, X., Chang, Z., and Chen, P.R. (2011). A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nat. Chem. Biol.* **7**, 671–677.
- Zückert, W.R. (2014). Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochim. Biophys. Acta* **1843**, 1509–1516.

Note Added in Proof

While this paper was under revision, Konovalova et al. reported the surface exposure of portions of RcsF via OM β -barrels.

Konovalova, A., Perlman, D.H., Cowles, C.E., and Silhavy, T.J. (2014). Transmembrane domain of surface-exposed outer membrane lipoprotein RcsF is threaded through the lumen of β -barrel proteins. *Proc. Natl. Acad. Sci. USA* **111**, E4350–E4358.

A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping

Suhas S.P. Rao,^{1,2,3,4,10} Miriam H. Huntley,^{1,2,3,4,5,10} Neva C. Durand,^{1,2,3,4} Elena K. Stamenova,^{1,2,3,4} Ivan D. Bochkov,^{1,2,3} James T. Robinson,^{1,4} Adrian L. Sanborn,^{1,2,3,6} Ido Machol,^{1,2,3} Arina D. Omer,^{1,2,3} Eric S. Lander,^{4,7,8,*} and Erez Lieberman Aiden^{1,2,3,4,9,*}

¹The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA

²Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

³Department of Computer Science, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005, USA

⁴Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA

⁵School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

⁶Department of Computer Science, Stanford University, Stanford, CA 94305, USA

⁷Department of Biology, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

⁸Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

⁹Center for Theoretical Biological Physics, Rice University, Houston, TX 77030, USA

¹⁰Co-first author

*Correspondence: lander@broadinstitute.org (E.S.L.), erez@erez.com (E.L.A.)

<http://dx.doi.org/10.1016/j.cell.2014.11.021>

SUMMARY

We use in situ Hi-C to probe the 3D architecture of genomes, constructing haploid and diploid maps of nine cell types. The densest, in human lymphoblastoid cells, contains 4.9 billion contacts, achieving 1 kb resolution. We find that genomes are partitioned into contact domains (median length, 185 kb), which are associated with distinct patterns of histone marks and segregate into six subcompartments. We identify ~10,000 loops. These loops frequently link promoters and enhancers, correlate with gene activation, and show conservation across cell types and species. Loop anchors typically occur at domain boundaries and bind CTCF. CTCF sites at loop anchors occur predominantly (>90%) in a convergent orientation, with the asymmetric motifs “facing” one another. The inactive X chromosome splits into two massive domains and contains large loops anchored at CTCF-binding repeats.

INTRODUCTION

The spatial organization of the human genome is known to play an important role in the transcriptional control of genes (Cremer and Cremer, 2001; Sexton et al., 2007; Bickmore, 2013). Yet important questions remain, like how distal regulatory elements, such as enhancers, affect promoters, and how insulators can abrogate these effects (Banerji et al., 1981; Blackwood and Kadonaga, 1998; Gaszner and Felsenfeld, 2006). Both phenomena are thought to involve the formation of protein-mediated “loops” that bring pairs of genomic sites that lie far apart along the linear genome into proximity (Schleif, 1992).

Various methods have emerged to assess the 3D architecture of the nucleus. In one seminal study, the binding of a protein to sites at opposite ends of a restriction fragment created a loop, which was detectable because it promoted the formation of DNA circles in the presence of ligase. Removal of the protein or either of its binding sites disrupted the loop, eliminating this “cyclization enhancement” (Mukherjee et al., 1988). Subsequent adaptations of cyclization enhancement made it possible to analyze chromatin folding in vivo, including nuclear ligation assay (Cullen et al., 1993) and chromosome conformation capture (Dekker et al., 2002), which analyze contacts made by a single locus, extensions such as 5C for examining several loci simultaneously (Dostie et al., 2006), and methods such as ChIA-PET for examining all loci bound by a specific protein (Fullwood et al., 2009).

To interrogate all loci at once, we developed Hi-C, which combines DNA proximity ligation with high-throughput sequencing in a genome-wide fashion (Lieberman-Aiden et al., 2009). We used Hi-C to demonstrate that the genome is partitioned into numerous domains that fall into two distinct compartments. Subsequent analyses have suggested the presence of smaller domains and have led to the important proposal that compartments are partitioned into condensed structures ~1 Mb in size, dubbed “topologically associated domains” (TADs) (Dixon et al., 2012; Nora et al., 2012). In principle, Hi-C could also be used to detect loops across the entire genome. To achieve this, however, extremely large data sets and rigorous computational methods are needed. Recent efforts have suggested that this is an increasingly plausible goal (Sexton et al., 2012; Jin et al., 2013).

Here, we report the results of an effort to comprehensively map chromatin contacts genome-wide, using in situ Hi-C, in which DNA-DNA proximity ligation is performed in intact nuclei. The protocol facilitates the generation of much denser Hi-C maps. The maps reported here comprise over 5 Tb of sequence

data recording over 15 billion distinct contacts, an order of magnitude larger than all published Hi-C data sets combined. Using these maps, we are able to clearly discern domain structure, compartmentalization, and thousands of chromatin loops. In addition to haploid maps, we were also able to create diploid maps analyzing each chromosomal homolog separately. The maps provide a picture of genomic architecture with resolution down to 1 kb.

RESULTS

In Situ Hi-C Methodology and Maps

Our in situ Hi-C protocol combines our original Hi-C protocol (here called dilution Hi-C) with nuclear ligation assay (Cullen et al., 1993), in which DNA is digested using a restriction enzyme, DNA-DNA proximity ligation is performed in intact nuclei, and the resulting ligation junctions are quantified. Our in situ Hi-C protocol involves crosslinking cells with formaldehyde, permeabilizing them with nuclei intact, digesting DNA with a suitable 4-cutter restriction enzyme (such as Mbol), filling the 5'-overhangs while incorporating a biotinylated nucleotide, ligating the resulting blunt-end fragments, shearing the DNA, capturing the biotinylated ligation junctions with streptavidin beads, and analyzing the resulting fragments with paired-end sequencing (Figure 1A). This protocol resembles a recently published single-cell Hi-C protocol (Nagano et al., 2013), which also performed DNA-DNA proximity ligation inside nuclei to study nuclear architecture in individual cells. Our updated protocol has three major advantages over dilution Hi-C. First, in situ ligation reduces the frequency of spurious contacts due to random ligation in dilute solution—as evidenced by a lower frequency of junctions between mitochondrial and nuclear DNA in the captured fragments and by the higher frequency of random ligations observed when the supernatant is sequenced (Extended Experimental Procedures available online). This is consistent with a recent study showing that ligation junctions formed in solution are far less meaningful (Gavrilov et al., 2013). Second, the protocol is faster, requiring 3 days instead of 7 (Extended Experimental Procedures). Third, it enables higher resolution and more efficient cutting of chromatinized DNA, for instance, through the use of a 4-cutter rather than a 6-cutter (Data S1, I).

A Hi-C map is a list of DNA-DNA contacts produced by a Hi-C experiment. By partitioning the linear genome into “loci” of fixed size (e.g., bins of 1 Mb or 1 kb), the Hi-C map can be represented as a “contact matrix” M , where the entry $M_{i,j}$ is the number of contacts observed between locus L_i and locus L_j . (A “contact” is a read pair that remains after we exclude reads that are duplicates, that correspond to unligated fragments, or that do not align uniquely to the genome.) The contact matrix can be visualized as a heatmap, whose entries we call “pixels.” An “interval” refers to a set of consecutive loci; the contacts between two intervals thus form a “rectangle” or “square” in the contact matrix. We define the “matrix resolution” of a Hi-C map as the locus size used to construct a particular contact matrix and the “map resolution” as the smallest locus size such that 80% of loci have at least 1,000 contacts. The map resolution is meant to reflect the finest scale at which one can reliably discern local features.

Contact Maps Spanning Nine Cell Lines Containing over 15 Billion Contacts

We constructed in situ Hi-C maps of nine cell lines in human and mouse (Table S1). Whereas our original Hi-C experiments had a map resolution of 1 Mb, these maps have a resolution of 1 kb or 5 kb. Our largest map, in human GM12878 B-lymphoblastoid cells, contains 4.9 billion pairwise contacts and has a map resolution of 950 bp (“kilobase resolution”) (Table S2). We also generated eight in situ Hi-C maps at 5 kb resolution, using cell lines representing all human germ layers (IMR90, HMEC, NHEK, K562, HUVEC, HeLa, and KBM7) as well as mouse B-lymphoblasts (CH12-LX) (Table S1). Each map contains between 395 M and 1.1 B contacts.

When we used our original dilution Hi-C protocol to generate maps of GM12878, IMR90, HMEC, NHEK, HUVEC, and CH12-LX, we found that, as expected, in situ Hi-C maps were superior at high resolutions, but closely resembled dilution Hi-C at lower resolutions. For instance, our dilution map of GM12878 (3.2 billion contacts) correlated highly with our in situ map at 500, 50, and 25 kb resolutions ($R > 0.96, 0.90$, and 0.87 , respectively) (Data S1, I; Figure S1).

We also performed 112 supplementary Hi-C experiments using three different protocols (in situ Hi-C, dilution Hi-C, and Tethered Conformation Capture) while varying a wide array of conditions such as extent of crosslinking, restriction enzyme, ligation volume/time, and biotinylated nucleotide. These include several in situ Hi-C experiments in which the formaldehyde crosslinking step was omitted, which demonstrate that the structural features we observe cannot be due to the crosslinking procedure. In total, 201 independent Hi-C experiments were successfully performed, many of which are presented in Data S1 and S2.

To account for nonuniformities in coverage due to the number of restriction sites at a locus or the accessibility of those sites to cutting (Lieberman-Aiden et al., 2009; Yaffe and Tanay, 2011) we use a matrix-balancing algorithm due to Knight and Ruiz (2012) (Extended Experimental Procedures).

Adequate tools for visualization of these large data sets are essential. We have therefore created the “Juicebox” visualization system that enables users to explore contact matrices, zoom in and out, compare Hi-C matrices to 1D tracks, superimpose all features reported in this paper onto the data, and contrast different Hi-C maps. All contact data and feature sets reported here can be explored interactively via Juicebox at <http://www.aidenlab.org/juicebox/>.

The Genome Is Partitioned into Small Domains Whose Median Length Is 185 kb

We began by probing the 3D partitioning of the genome. In our earlier experiments at 1 Mb map resolution (Lieberman-Aiden et al., 2009), we saw large squares of enhanced contact frequency tiling the diagonal of the contact matrices. These squares partitioned the genome into 5–20 Mb intervals, which we call “megadomains.”

We also found that individual 1 Mb loci could be assigned to one of two long-range contact patterns, which we called compartments A and B, with loci in the same compartment showing more frequent interaction. Megadomains—and the associated squares along the diagonal—arise when all of the 1 Mb loci in



(B) Contact matrices from chromosome 14: the whole chromosome, at 500 kb resolution (top); 86–96 Mb/50 kb resolution (middle); 94–95 Mb/5 kb resolution (bottom). Left: GM12878, primary experiment; Right: biological replicate. The 1D regions corresponding to a contact matrix are indicated in the diagrams above and at left. The intensity of each pixel represents the normalized number of contacts between a pair of loci. Maximum intensity is indicated in the lower left of each panel.

(C) We compare our map of chromosome 7 in GM12878 (last column) to earlier Hi-C maps: Lieberman-Aiden et al. (2009), Kalhor et al. (2012), and Jin et al. (2013). (D) Overview of features revealed by our Hi-C maps. Top: the long-range contact pattern of a locus (left) indicates its nuclear neighborhood (right). We detect at least six subcompartments, each bearing a distinctive pattern of epigenetic features. Middle: squares of enhanced contact frequency along the diagonal (left) indicate the presence of small domains of condensed chromatin, whose median length is 185 kb (right). Bottom: peaks in the contact map (left) indicate the presence of loops (right). These loops tend to lie at domain boundaries and bind CTCF in a convergent orientation.

Cell 159, 1665–1680, December 18, 2014 ©2014 Elsevier Inc. 1667

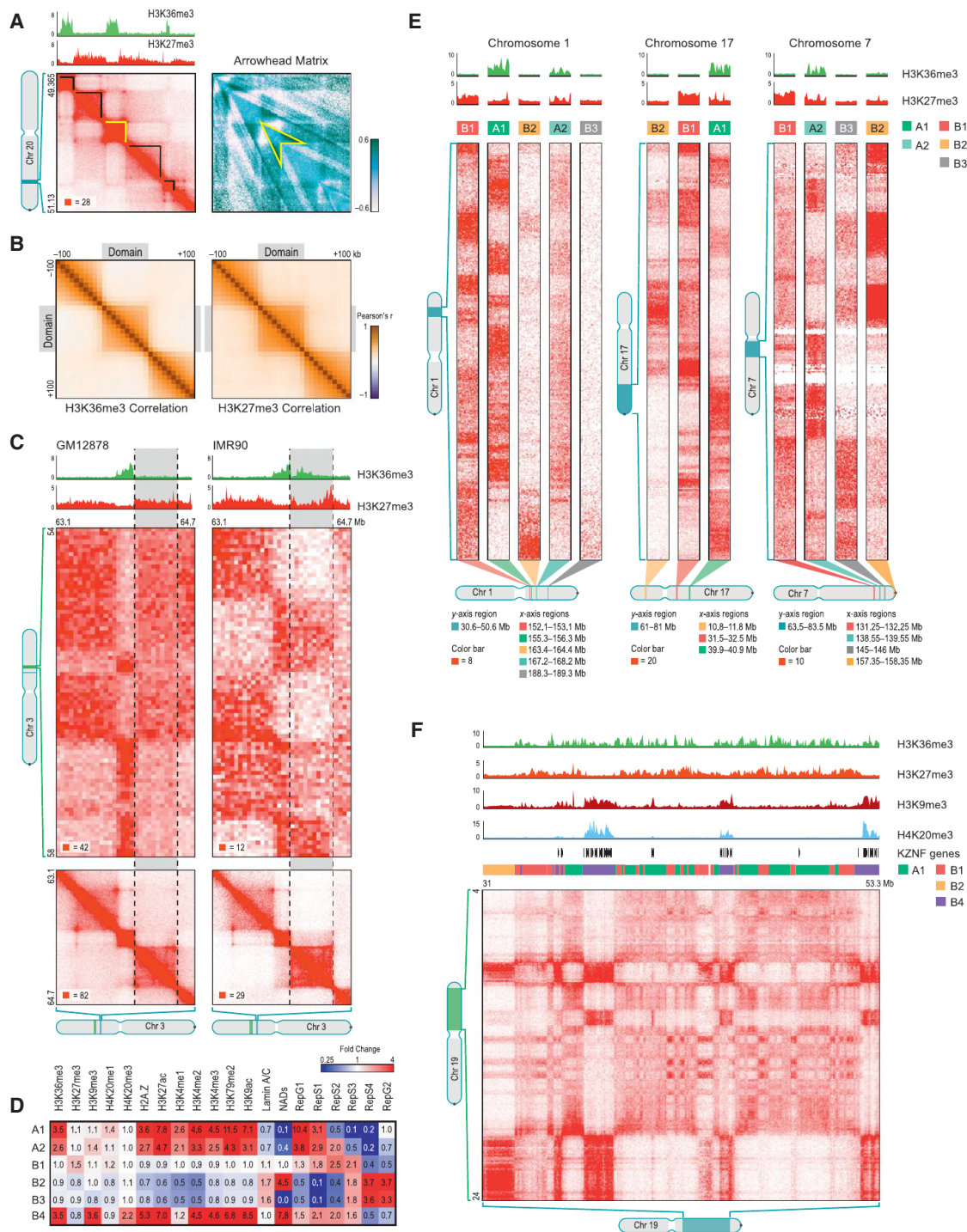


Figure 2. The Genome Is Partitioned into Contact Domains that Segregate into Nuclear Subcompartments Corresponding to Different Patterns of Histone Modifications

(A) We annotate thousands of domains across the genome (left, black highlight). To do so, we define an arrowhead matrix A (right) such that $A_{i,i+d} = (M^*_{i,i-d} - M^*_{i,i+d}) / (M^*_{i,i-d} + M^*_{i,i+d})$, where M^* is the normalized contact matrix. This transformation replaces domains with an arrowhead-shaped motif pointing toward the domain's upper-left corner (example in yellow); we identify these arrowheads using dynamic programming. See Experimental Procedures.

(B) Pearson correlation matrices of the histone mark signal between pairs of loci inside and within 100 kb of a domain. Left: H3K36me3; Right: H3K27me3.

(C) Conserved contact domains on chromosome 3 in GM12878 (left) and IMR90 (right). In GM12878, the highlighted domain (gray) is enriched for H3K27me3 and depleted for H3K36me3. In IMR90, the situation is reversed. Marks at flanking domains are the same in both: the domain to the left is enriched for H3K36me3 and the domain to the right is enriched for H3K27me3. The flanking domains have long-range contact patterns that differ from one another and are preserved in both

(legend continued on next page)

an interval exhibit the same genome-wide contact pattern. Compartment A is highly enriched for open chromatin; compartment B is enriched for closed chromatin (Lieberman-Aiden et al., 2009; Kalhor et al., 2012; Sexton et al., 2012).

In our new, higher resolution maps (200- to 1,000-fold more contacts), we observe many small squares of enhanced contact frequency that tile the diagonal of each contact matrix (Figure 2A). We used the Arrowhead algorithm (see Experimental Procedures) to annotate these contact domains genome-wide. The observed domains ranged in size from 40 kb to 3 Mb (median size 185 kb). As with megadomains, there is an abrupt drop in contact frequency (33%) for pairs of loci on opposite sides of the domain boundary (Figure S2G). Contact domains are often preserved across cell types (Figures S3A and S3B).

The presence of smaller domains in Hi-C maps is consistent with several other recent studies (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). We explore the relationship between the domains we annotate and those annotated in prior studies in the Discussion.

Contact Domains Exhibit Consistent Histone Marks Whose Changes Are Associated with Changes in Long-Range Contact Pattern

Loci within a contact domain show correlated histone modifications for eight different factors (H3K36me3, H3K27me3, H3K4me1, H3K4me2, H3K4me3, H3K9me3, H3K79me2, and H4K20me1) based on data from the ENCODE project in GM12878 cells (ENCODE Project Consortium, 2012). By contrast, loci at comparable distance but residing in different domains showed much less correlation in chromatin state (Figures 2B, S2I, and S2K; Extended Experimental Procedures). Strikingly, changes in a domain's chromatin state are often accompanied by changes in the long-range contact pattern of domain loci (i.e., the pattern of contacts between loci in the domain and other loci genome-wide), indicating that changes in chromatin pattern are accompanied by shifts in a domain's nuclear neighborhood (Figures 2C and S3C–S3E; Extended Experimental Procedures). This observation is consistent with microscopy studies associating changes in gene expression with changes in nuclear localization (Finlan et al., 2008).

There Are at Least Six Nuclear Subcompartments with Distinct Patterns of Histone Modifications

Next, we partitioned loci into categories based on long-range contact patterns alone, using four independent approaches: manual annotation and three unsupervised clustering algorithms (HMM, K-means, Hierarchical). All gave similar results (Figure S4B; Extended Experimental Procedures). We then investigated the biological meaning of these categories.

When we analyzed the data at low matrix resolution (1 Mb), we reproduced our earlier finding of two compartments (A and B). At high resolution (25 kb), we found evidence for at least five “subcompartments” defined by their long-range interaction patterns, both within and between chromosomes. These findings expand on earlier reports suggesting three compartments in human cells (Yaffe and Tanay, 2011). We found that the median length of an interval lying completely within a subcompartment is 300 kb. Although the subcompartments are defined solely based on their Hi-C interaction patterns, they exhibit distinct genomic and epigenomic content.

Two of the five interaction patterns are correlated with loci in compartment A (Figure S4E). We label the loci exhibiting these patterns as belonging to subcompartments A1 and A2. Both A1 and A2 are gene dense, have highly expressed genes, harbor activating chromatin marks such as H3K36me3, H3K79me2, H3K27ac, and H3K4me1 and are depleted at the nuclear lamina and at nucleolus-associated domains (NADs) (Figures 2D, 2E, and S4I; Table S3). While both A1 and A2 exhibit early replication times, A1 finishes replicating at the beginning of S phase, whereas A2 continues replicating into the middle of S phase. A2 is more strongly associated with the presence of H3K9me3 than A1, has lower GC content, and contains longer genes (2.4-fold).

The other three interaction patterns (labeled B1, B2, and B3) are correlated with loci in compartment B (Figure S4E) and show very different properties. Subcompartment B1 correlates positively with H3K27me3 and negatively with H3K36me3, suggestive of facultative heterochromatin (Figures 2D and 2E). Replication of this subcompartment peaks during the middle of S phase. Subcompartments B2 and B3 tend to lack all of the above-noted marks and do not replicate until the end of S phase (see Figure 2D). Subcompartment B2 includes 62% of pericentromeric heterochromatin (3.8-fold enrichment) and is enriched at the nuclear lamina (1.8-fold) and at NADs (4.6-fold). Subcompartment B3 is enriched at the nuclear lamina (1.6-fold), but strongly depleted at NADs (76-fold).

Upon closer visual examination, we noticed the presence of a sixth pattern on chromosome 19 (Figure 2F). Our genome-wide clustering algorithm missed this pattern because it spans only 11 Mb, or 0.3% of the genome. When we repeated the algorithm on chromosome 19 alone, the additional pattern was detected. Because this sixth pattern correlates with the Compartment B pattern, we labeled it B4. Subcompartment B4 comprises a handful of regions, each of which contains many KRAB-ZNF superfamily genes. (B4 contains 130 of the 278 KRAB-ZNF genes in the genome, a 65-fold enrichment). As noted in previous studies (Vogel et al., 2006; Hahn et al., 2011), these regions exhibit a highly distinctive chromatin pattern, with strong enrichment for

cell types. In IMR90, the highlighted domain is marked by H3K36me3 and its long-range contact pattern matches the similarly-marked domain on the left. In GM12878, it is decorated with H3K27me3, and the long-range pattern switches, matching the similarly-marked domain to the right. Diagonal submatrices, 10 kb resolution; long-range interaction matrices, 50 kb resolution.

(D) Each of the six long-range contact patterns we observe exhibits a distinct epigenetic profile (data sources are listed in Table S3). Each subcompartment also has a visually distinctive contact pattern.

(E) Each example shows part of the long-range contact patterns for several nearby genomic intervals lying in different subcompartments.

(F) A large contiguous region on chromosome 19 contains intervals in subcompartments A1, B1, B2, and B4.

See also Figures S2, S3, and S4 and Data S1, III–IV.

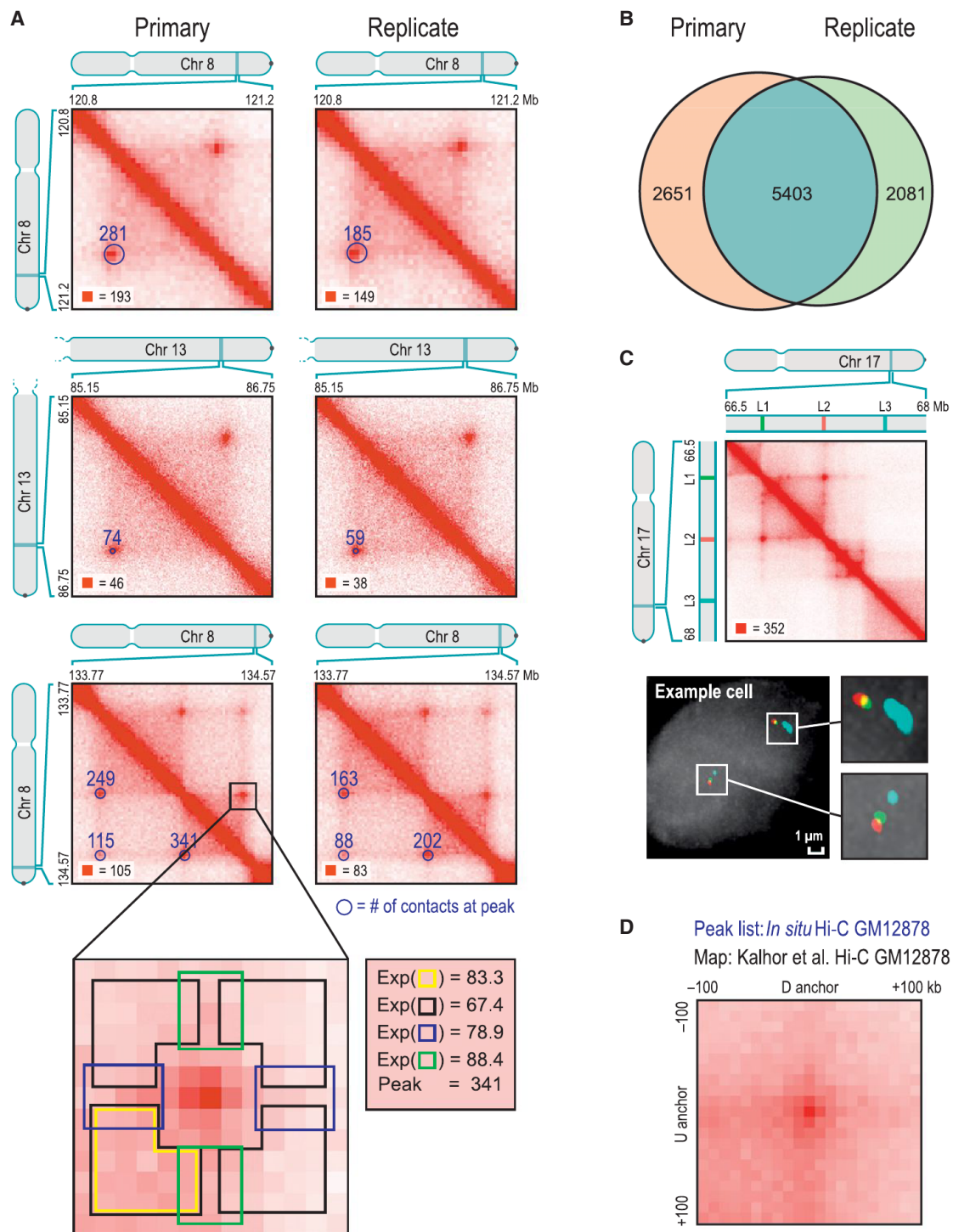


Figure 3. We Identify Thousands of Chromatin Loops Genome-wide Using a Local Background Model

(A) We identify peaks by detecting pixels that are enriched with respect to four local neighborhoods (blowout): horizontal (blue), vertical (green), lower-left (yellow), and donut (black). These “peak” pixels indicate the presence of a loop and are marked with blue circles (radius = 20 kb) in the lower-left of each heatmap. The number of raw contacts at each peak is indicated. Left: primary GM12878 map; Right: replicate; annotations are completely independent. All contact matrices in this and subsequent figures are 10 kb resolution unless noted.

(B) Overlap in peak annotations between replicates.

(C) Top: location of 3D-FISH probes used to verify a peak in the chromosome 17 contact map. Bottom: example cell.

(legend continued on next page)

both activating chromatin marks, such as H3K36me3, and heterochromatin-associated marks, such as H3K9me3 and H4K20me3.

Approximately 10,000 Peaks Mark the Position of Chromatin Loops

We next sought to identify the positions of chromatin loops by using an algorithm to search for pairs of loci that show significantly closer proximity with one another than with the loci lying between them (Figure 3A). Such pairs correspond to pixels with higher contact frequency than typical pixels in their neighborhood. We refer to these pixels as “peaks” in the Hi-C contact matrix and to the corresponding pair of loci as “peak loci.” Peaks reflect the presence of chromatin loops, with the peak loci being the anchor points of the chromatin loop. (Because contact frequencies vary across the genome, we define peak pixels relative to the local background. We note that some papers [Sanyal et al., 2012; Jin et al., 2013] have sought to define peaks relative to a genome-wide average. This choice is problematic because, for example, many pixels within a domain may be reported as peaks despite showing no locally distinctive proximity; see Discussion.)

Our algorithm detected 9,448 peaks in the *in situ* Hi-C map for GM12878 at 5 kb matrix resolution. These peaks are associated with a total of 12,903 distinct peak loci (some peak loci are associated with more than one peak). The vast majority of peaks (98%) reflected loops between loci that are <2 Mb apart.

These findings were reproducible across all of our high-resolution Hi-C maps. Examining the primary and replicate maps separately, we found 8,054 peaks in the former and 7,484 peaks in the latter, with 5,403 in both lists (see Figures 3A and 3B; Data S1, V; Table S4). The differences were almost always the result of our conservative peak-calling criteria (Extended Experimental Procedures). We also called peaks using our GM12878 dilution Hi-C experiment. Because the map is sparser and thus noisier, we called only 3,073 peaks. Nonetheless, 65% of these peaks were also present in the list of peaks from our *in situ* Hi-C data set, again reflecting high interreplicate reproducibility.

To independently confirm that peak loci are closer than neighboring locus pairs, we performed 3D-FISH (Beliveau et al., 2012) on four loops (Table S5). In each case, we compared two peak loci, *L1* and *L2*, with a control locus, *L3*, that lies an equal genomic distance away from *L2* but on the opposite side (Figures 3C and S5B). In all cases, the 3D-distance between *L1* and *L2* was consistently shorter than the 3D-distance between *L2* and *L3* (Extended Experimental Procedures).

We also confirmed that our list of peaks was consistent with previously published Hi-C maps. Although earlier maps contained too few contacts to reliably call individual peaks, we developed a method called Aggregate Peak Analysis (APA) that compares the aggregate enrichment of our peak set in these low-resolution maps to the enrichment seen when our peak set is translated in any direction (Experimental Procedures). APA

showed strong consistency between our loop calls and all six previously published Hi-C experiments in lymphoblastoid cell lines (Lieberman-Aiden et al., 2009; Kalhor et al., 2012) (Figure 3D; Data S2, I.E; Table S6).

Finally, we demonstrated that the peaks observed were robust to particular protocol conditions by performing APA on our GM12878 dilution Hi-C map and on our 112 supplemental Hi-C experiments exploring a wide range of protocol variants. Enrichment was seen in every experiment. Notably, these include five experiments (HIC043-HIC047; Table S1) in which the Hi-C protocol was performed without crosslinking, demonstrating that the peaks observed in our experiments cannot be byproducts of the formaldehyde-crosslinking procedure.

Conservation of Peaks among Human Cell Lines and across Evolution

We also identified peaks in the other seven human cell lines (Table S1). Because these maps contain fewer contacts, sensitivity is reduced, and fewer peaks are observed (ranging from 2,634 to 8,040). APA confirmed that these peak calls were consistent with the dilution Hi-C maps reported here (in IMR90, HMEC, HUVEC, and NHEK), as well as with all previously published Hi-C maps in these cell types (Lieberman-Aiden et al., 2009; Dixon et al., 2012; Jin et al., 2013) (Data S2, I.F).

We found that peaks were often conserved across cell types (Figure 4A): between 55% and 75% of the peaks found in any given cell type were also found in GM12878 (Figure S5D).

Next, we compared peaks across species. In CH12-LX mouse B-lymphoblasts, we identified 2,927 high-confidence contact domains and 3,331 peaks. When we examined orthologous regions in GM12878, we found that 50% of peaks and 45% of domains called in mouse were also called in humans. This suggests substantial conservation of 3D genome structure across the mammals (Figures 4B–4E).

Loops Anchored at a Promoter Are Associated with Enhancers and Increased Gene Activation

Various lines of evidence indicate that many of the observed loops are associated with gene regulation.

First, our peaks frequently have a known promoter at one peak locus (as annotated by ENCODE's ChromHMM) (Hoffman et al., 2013) and a known enhancer at the other (Figure 5A). For instance, 2,854 of the 9,448 peaks in our GM12878 map bring together known promoters and known enhancers (30% versus 7% expected by chance). The peaks include classic promoter-enhancer loops, such as at *MYC* (chr8:128.35–128.75 Mb, in HMEC) and alpha-globin (chr16:0.15–0.22 Mb, in K562). Second, genes whose promoters are associated with a loop are much more highly expressed than genes whose promoters are not associated with a loop (6-fold).

Third, the presence of cell type-specific peaks is associated with changes in expression. When we examined RNA sequencing (RNA-seq) data produced by ENCODE, we found

(D) APA plot shows the aggregate signal from the 9,448 GM12878 loops we report by summing submatrices surrounding each peak in a low-resolution GM12878 Hi-C map due to Kalhor et al. (2012). Although individual peaks cannot be seen in the Kalhor et al. (2012) data (that contains 42 M contacts), the peak at the center of the APA plot indicates that the aggregate signal from our peak set as a whole can be clearly discerned using their data set. See also Figure S5, Data S1, V. and Data S2, I, and Tables S4, S5, and S6.

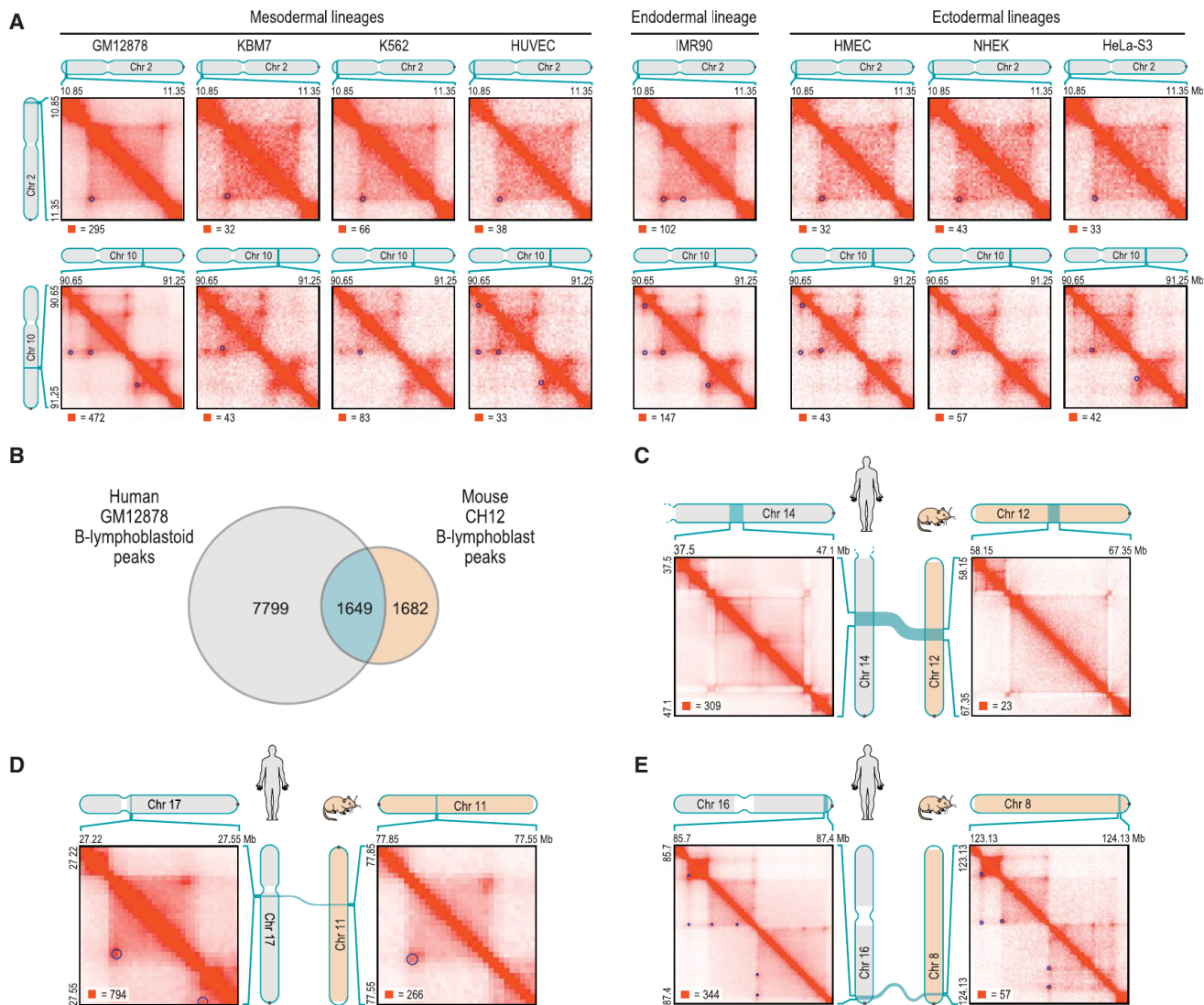


Figure 4. Loops Are Often Preserved across Cell Types and from Human to Mouse

(A) Examples of peak and domain preservation across cell types. Annotated peaks are circled in blue. All annotations are completely independent.

(B) Of the 3,331 loops we annotate in mouse CH12-LX, 1,649 (50%) are orthologous to loops in human GM12878.

(C–E) Conservation of 3D structure in syntenic blocks. The contact matrices in (C) are shown at 25 kb resolution. (D) and (E) are shown at 10 kb resolution.

that the appearance of a loop in a cell type was frequently accompanied by the activation of a gene whose promoter overlapped one of the peak loci. For example, a cell-type-specific loop is anchored at the promoter of the gene encoding L-selectin (*SELL*), which is expressed in GM12878 (where the loop is present), but not in IMR90 (where the loop is absent, Figure 5B). Genome-wide, we observed 557 loops in GM12878 that were clearly absent in IMR90. The corresponding peak loci overlapped the promoters of 43 genes that were markedly upregulated (>50-fold) in GM12878, but of only one gene that was markedly upregulated in IMR90. Conversely, we found 510 loops in IMR90 that were clearly absent in GM12878. The corresponding peak loci overlapped the promoters of 94 genes that were markedly upregulated in IMR90, but of only three genes that were

markedly upregulated in GM12878. When we compared GM12878 to the five other human cell types for which ENCODE RNA-seq data were available, the results were very similar (Figure 5C; Table S7).

Occasionally, gene activation is accompanied by the emergence of a cell-type-specific network of peaks. Figure 5D illustrates the case of *ADAMTS1*, which encodes a protein involved in fibroblast migration. The gene is expressed in IMR90, where its promoter is involved in six loops. In GM12878, it is not expressed, and the promoter is involved in only two loops. Many of the IMR90 peak loci form transitive peaks with one another (see discussion of “transitivity” below), suggesting that the *ADAMTS1* promoter and the six distal sites may all be located at a single spatial hub.

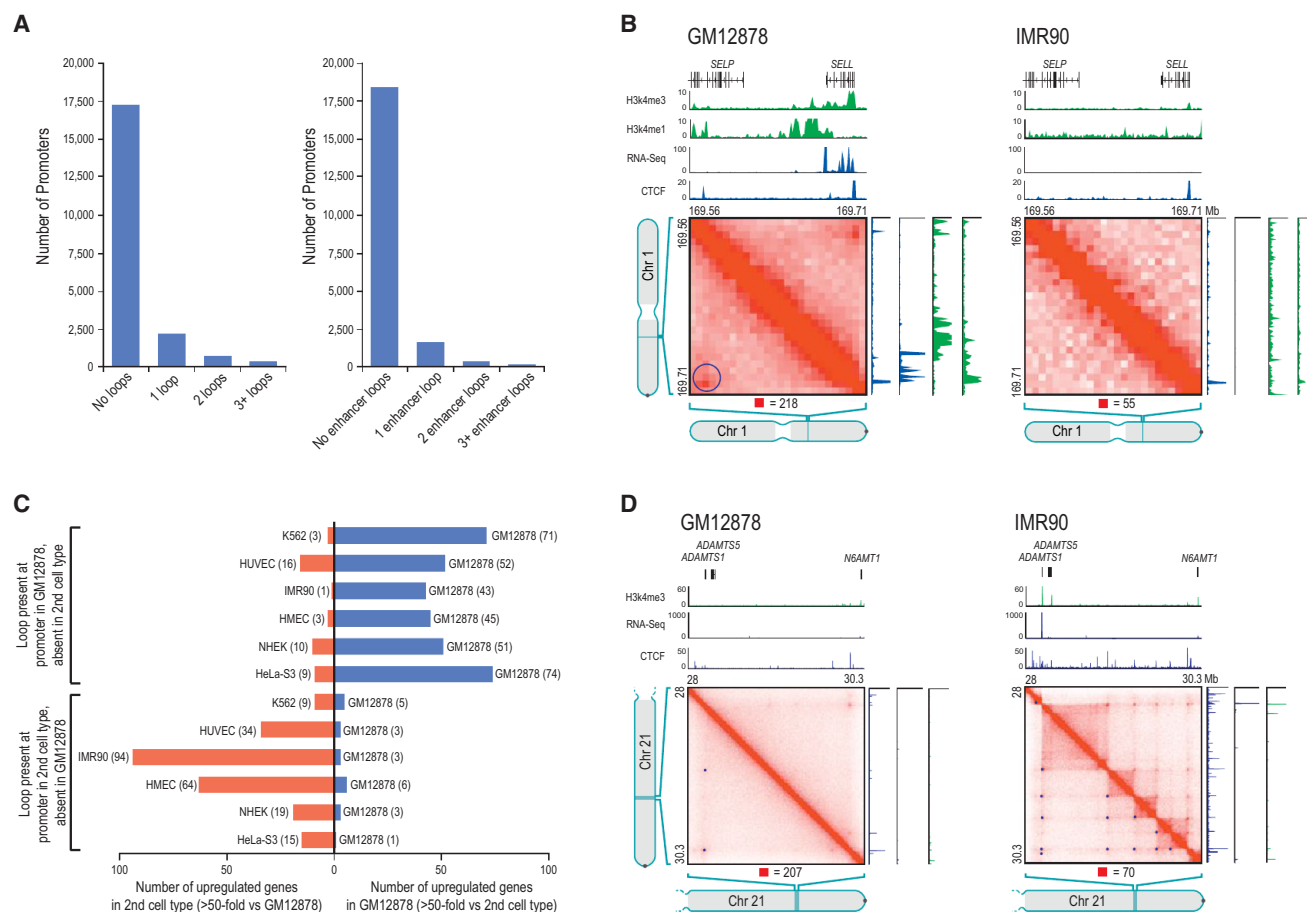


Figure 5. Loops between Promoters and Enhancers Are Strongly Associated with Gene Activation

(A) Histogram showing loop count at promoters (left); restricted to loops where the distal peak locus contains an enhancer (right).

(B) Left: a loop in GM12878, with one anchor at the *SELL* promoter and the other at a distal enhancer. The gene is on. Right: the loop is absent in IMR90, where the gene is off.

(C) Genes whose promoters participate in a loop in GM12878 but not in a second cell type are frequently upregulated in GM12878 and vice versa.

(D) Left: two loops in GM12878 are anchored at the promoter of the inactive *ADAMTS1* gene. Right: a series of loops and domains appear, along with transitive looping. *ADAMTS1* is on.

See also Data S1, VI and Table S7.

These observations are consistent with the classic model in which looping between a promoter and enhancer activates a target gene (Tolhuis et al., 2002; Amano et al., 2009; Ahmadiyeh et al., 2010).

Loops Frequently Delineate the Boundaries of Contact Domains

A large fraction of peaks (38%) coincide with the corners of a contact domain—that is, the peak loci are located at domain boundaries (Figures 6A and S6). Conversely, a large fraction of domains (39%) had peaks in their corner. Moreover, the appearance of a loop is usually (in 65% of cases) associated with the appearance of a domain demarcated by the loop. Because this configuration is so common, we use the term “loop domain” to refer to contact domains whose endpoints form a chromatin loop.

In some cases, adjacent loop domains (bounded by peak loci *L1-L2* and *L2-L3*, respectively) exhibit transitivity—that is, *L1* and

L3 also correspond to a peak. This may indicate that the three loci simultaneously colocalize at a single spatial position. However, many peaks do not exhibit transitivity, suggesting that the corresponding loci do not colocalize. Figure 6B shows a region on chromosome 4 exhibiting both configurations.

We also found that overlapping loops are strongly disfavored: pairs of loops *L1-L3* and *L2-L4* (where *L1*, *L2*, *L3* and *L4* occur consecutively in the genome) are found 4-fold less often than expected under a random model (Extended Experimental Procedures).

The Vast Majority of Loops Are Associated with Pairs of CTCF Motifs in a Convergent Orientation

We next wondered whether peaks are associated with specific proteins. We examined the results of 86 chromatin immunoprecipitation sequencing (ChIP-seq) experiments performed by ENCODE in GM12878. We found that the vast majority of peak

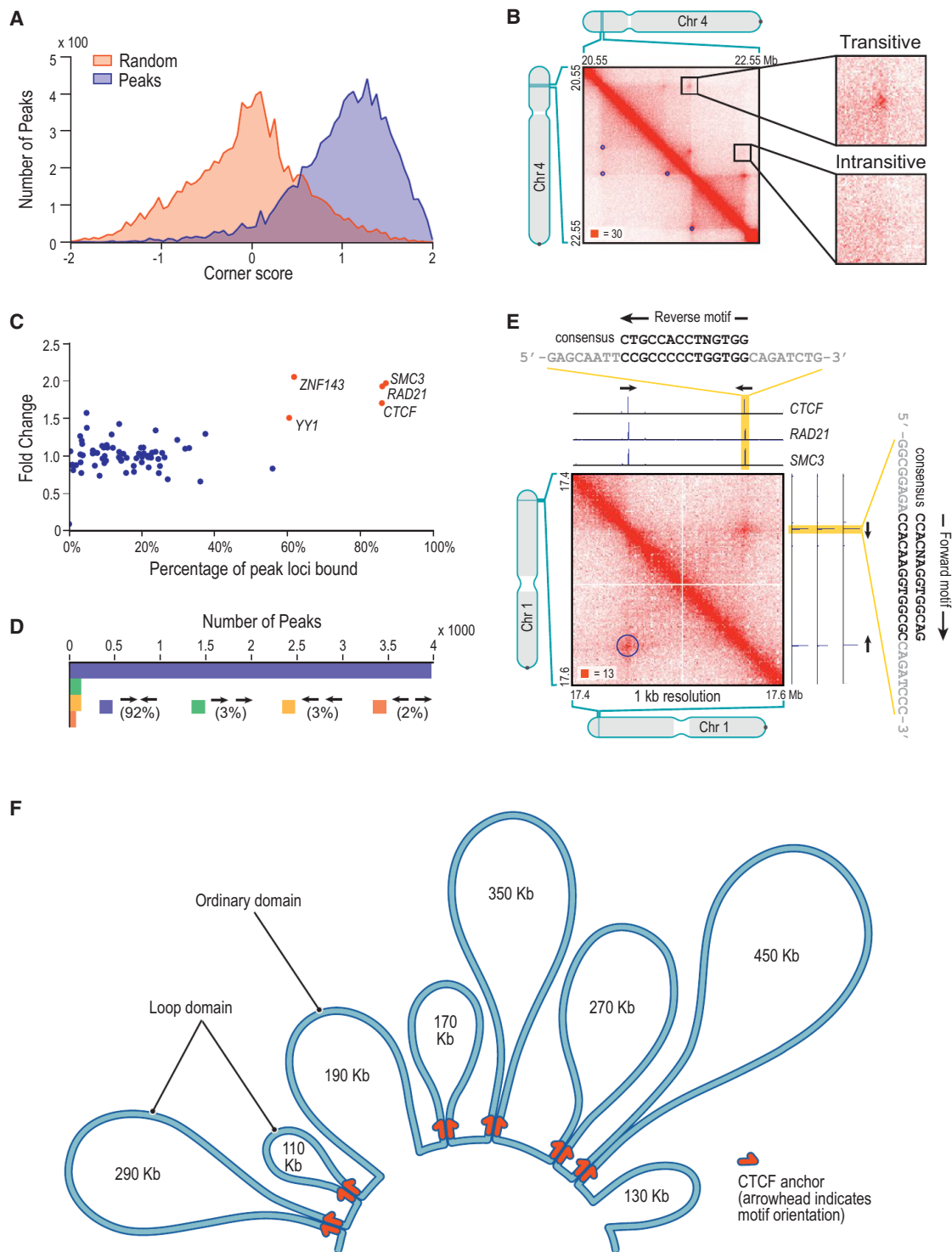


Figure 6. Many Loops Demarcate Contact Domains; The Vast Majority of Loops Are Anchored at a Pair of Convergent CTCF/RAD21/SMC3 Binding Sites

(A) Histograms of corner scores for peak pixels versus random pixels with an identical distance distribution.

(B) Contact matrix for chr4:20.55 Mb–22.55 Mb in GM12878, showing examples of transitive and intransitive looping behavior.

(C) Percent of peak loci bound versus fold enrichment for 76 DNA-binding proteins.

(D) The pairs of CTCF motifs that anchor a loop are nearly all found in the convergent orientation.

(legend continued on next page)

loci are bound by the insulator protein CTCF (86%) and the cohesin subunits RAD21 (86%) and SMC3 (87%) (Figure 6C). This is consistent with numerous reports, using a variety of experimental modalities, that suggest a role for CTCF and cohesin in mediating DNA loops (Splinter et al., 2006; Hou et al., 2008; Phillips and Corces, 2009). Because many of our loops demarcate domains, this observation is also consistent with studies suggesting that CTCF delimits structural and regulatory domains (Xie et al., 2007; Cuddapah et al., 2009; Dixon et al., 2012).

We found that most peak loci encompass a unique DNA site containing a CTCF-binding motif, to which all three proteins (CTCF, SMC3, and RAD21) were bound (5-fold enrichment). We were thus able to associate most of the peak loci (6,991 of 12,903, or 54%) with a specific CTCF-motif “anchor.”

The consensus DNA sequence for CTCF-binding sites is typically written as 5'-CCACNAGGTGGCAG-3'. Because the sequence is not palindromic, each CTCF motif has an orientation; we designate the consensus motif above as the “forward” orientation. Thus, a pair of CTCF sites on the same chromosome can have four possible orientations: (1) same direction on one strand, (2) same direction on the other strand, (3) convergent on opposite strands, and (4) divergent on opposite strands.

If CTCF sites were randomly oriented, one would expect all four orientations to occur equally often. But when we examined the 4,322 peaks in GM12878 where the two corresponding peak loci each contained a single CTCF-binding motif, we found that the vast majority (92%) of motif pairs are convergent (Figures 6D and 6E). Overall, the presence, at pairs of peak loci, of bound CTCF sites in the convergent orientation was enriched 102-fold over random expectation (Extended Experimental Procedures). The convergent orientation was overwhelmingly more frequent than the divergent orientation, despite the fact that divergent motifs also lie on opposing strands: in GM12878, the counts were 3,971-78 (51-fold enrichment, convergent versus divergent); in IMR90, 1,456-5 (291-fold); in HMEC, 968-11 (88-fold); in K562, 723-2 (362-fold); in HUVEC, 671-4 (168-fold); in HeLa, 301-3 (100-fold); in NHEK, 556-9 (62-fold); and in CH12-LX, 625-8 (78-fold). This pattern suggests that a pair of CTCF sites in the convergent orientation is required for the formation of a loop.

The observation that looped CTCF sites occur in the convergent orientation also allows us to analyze peak loci containing multiple CTCF-bound motifs to predict which motif instance plays a role in a given loop. In this way, we can associate nearly two-thirds of peak loci (8,175 of 12,903, or 63.4%) with a single CTCF-binding motif.

The specific orientation of CTCF sites at observed peaks provides evidence that our peak calls are biologically correct. Because randomly chosen CTCF pairs would exhibit each of the four orientations with equal probability, the near-perfect as-

sociation between our loop calls and the convergent orientation could not occur by chance ($p < 10^{-1,900}$, binomial distribution).

In addition, the presence of CTCF and RAD21 sites at many of our peaks provides an opportunity to compare our results to three recent ChIA-PET experiments reported by the ENCODE Consortium (in GM12878 and K562) in which ligation junctions bound to CTCF (or RAD21) were isolated and analyzed. We found strong concordance with our results in all three cases (Li et al., 2012; Heidari et al., 2014) (Extended Experimental Procedures).

The CTCF-Binding Exapted SINEB2 Repeat in Mouse Shows Preferential Orientation with Respect to Loops

In mouse, we found that 7% of peak anchors lie within SINEB2 repeat elements containing a CTCF motif, which has been exapted to be functional. (The spread of CTCF binding via retrotransposition of this element, which contains a CTCF motif in its consensus sequence, has been documented in prior studies [Bourque et al., 2008; Schmidt et al., 2012].) The CTCF motifs at peak anchors in SINEB2 elements show the same strong bias toward convergent orientation seen throughout the genome (89% are oriented toward the opposing loop anchor versus 94% genome-wide). The orientation of these CTCF motifs is aligned with the orientation of the SINEB2 consensus sequence in 97% of cases. This suggests that exaptation of a CTCF in a SINEB2 element is more likely when the orientation of the inserted SINEB2 is compatible with local loop structure.

Diploid Hi-C Maps Reveal Homolog-Specific Features, Including Imprinting-Specific Loops and Massive Domains and Loops on the Inactive X Chromosome

Because many of our reads overlap SNPs, it is possible to use GM12878 phasing data (McKenna et al., 2010; 1000 Genomes Project Consortium et al., 2012) to assign contacts to specific chromosomal homologs (Figure 7A; Table S8). Using these assignments, we constructed a “diploid” Hi-C map of GM12878 comprising both maternal (238 M contacts) and paternal (240 M) maps.

For autosomes, the maternal and paternal homologs exhibit very similar inter- and intrachromosomal contact profiles (Pearson's $R > 0.998$). One interchromosomal difference was notable: an elevated contact frequency between the paternal homologs of chromosome 6 and 11 that is consistent with an unbalanced translocation fusing chr11q:73.5 Mb and all distal loci (a stretch of over 60 Mb) to the telomere of chromosome 6p (Figures 7B and S7B). The signal intensity suggests that the translocation is present in between 1.2% and 5.6% of our cells (Extended Experimental Procedures). We tested this prediction by karyotyping 100 GM12878 cells using Giemsa staining and found three abnormal chromosomes, each showing the predicted

(E) A peak on chromosome 1 and corresponding ChIP-seq tracks. Both peak loci contain a single site bound by CTCF, RAD21, and SMC3. The CTCF motifs at the anchors exhibit a convergent orientation.

(F) A schematic rendering of a 2.1 Mb region on chromosome 20 (48.78–50.88 Mb). Eight domains tile the region, ranging in size from 110 kb to 450 kb; 95% of the region is contained inside a domain (contour lengths are shown to scale). Six of the eight domains are demarcated by loops between convergent CTCF-binding sites located at the domain boundaries. The other two domains are not demarcated by loops. The motif orientation is indicated by the direction of the arrow. Note that not every CTCF-binding site is shown.

See also Figure S6.

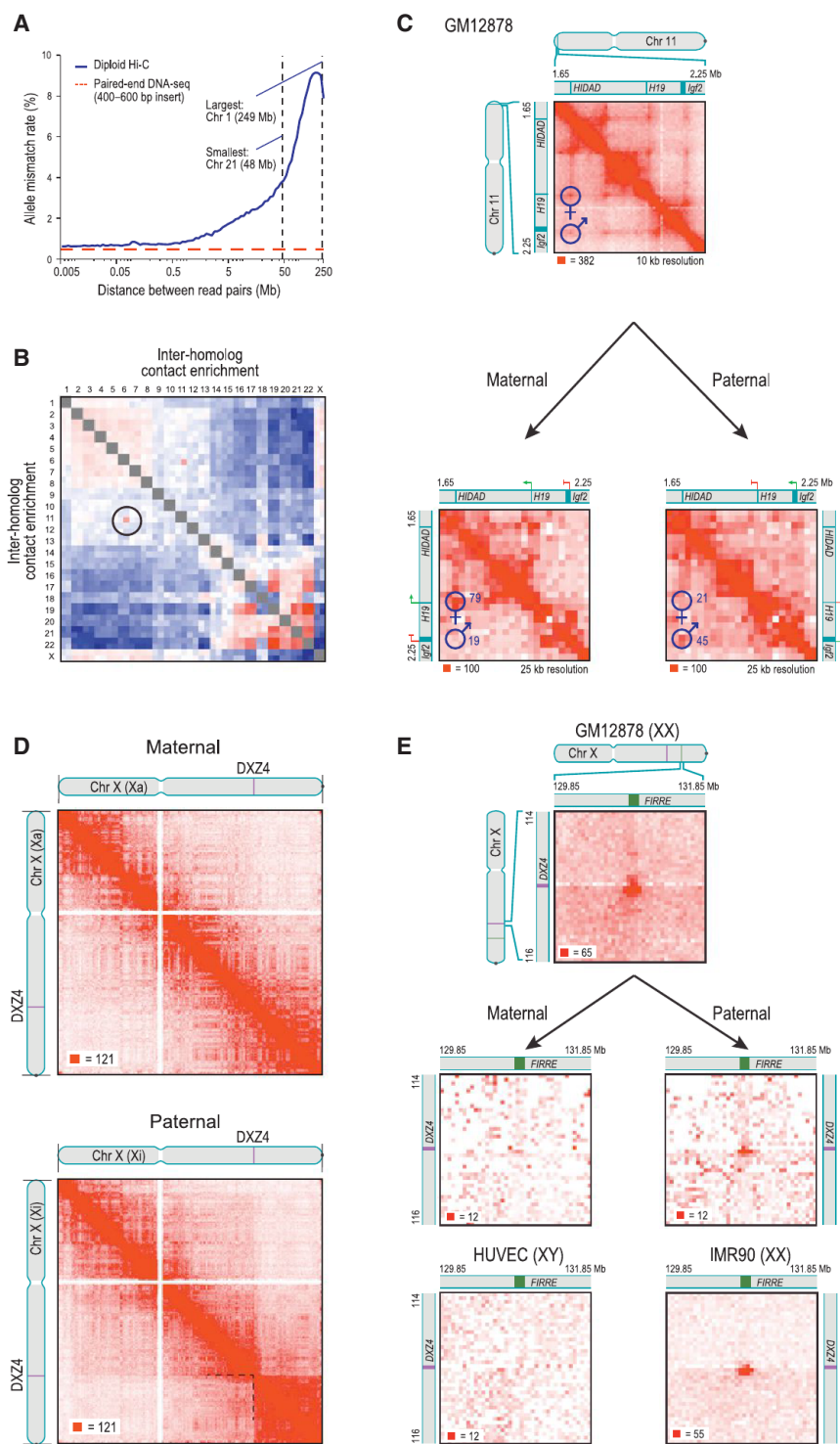


Figure 7. Diploid Hi-C Maps Reveal Superdomains and Superloops Anchored at CTCF-Binding Tandem Repeats on the Inactive X Chromosome

(A) The frequency of mismatch (maternal-paternal) in SNP allele assignment versus distance between two paired read alignments. Intrachromosomal read pairs are overwhelmingly intramolecular.

(B) Preferential interactions between homologs. Left/top is maternal; right/bottom is paternal. The aberrant contact frequency between 6/paternal and 11/paternal (circle) reveals a translocation.

(C) Top: in our unphased Hi-C map of GM12878, we observe two loops joining both the promoter of the maternally-expressed *H19* and the promoter of the paternally-expressed *Igf2* to a distal locus, HIDAD. Using diploid Hi-C maps, we phase these loops: the HIDAD-*H19* loop is present only on the maternal homolog (left) and the HIDAD-*Igf2* loop is present only on the paternal homolog (right).

(D) The inactive (paternal) copy of chromosome X (bottom) is partitioned into two massive “superdomains” not seen in the active (maternal) copy (top). *DXZ4* lies at the boundary. Contact matrices are shown at 500 kb resolution.

(E) The “superloop” between *FIRRE* and *DXZ4* is present in the unphased GM12878 map (top), in the paternal GM12878 map (middle right), and in the map of the female cell line IMR90 (bottom right); it is absent from the maternal GM12878 map (middle left) and the map of the male HUVEC cell line (bottom left). Contact matrices are shown at 50 kb resolution.

See also Figure S7 and Table S8.

of genomic imprinting. In our unphased maps, we clearly see two loops from a single distal locus at 1.72 Mb (that binds CTCF in the forward orientation) to loci located near the promoters of both *H19* and *Igf2* (both of which bind CTCF in the reverse orientation, i.e., the above consensus motif lies on the opposite strand; see Figure 7C). We refer to this distal locus as the *H19/Igf2* Distal Anchor Domain (HIDAD). Our diploid maps reveal that the loop to the *H19* region is present on the maternal chromosome (from which *H19* is expressed), but the loop to the *Igf2* region is absent or greatly attenuated. The opposite pattern is found on the paternal chromosome (from which *Igf2* is expressed).

Pronounced differences were seen on the diploid intrachromosomal maps of

chromosome X. The paternal X chromosome, which is usually inactive in GM12878, is partitioned into two massive domains (0–115 Mb and 115–155.3 Mb). These “superdomains” are not seen in the active, maternal X (Figure 7D). When we examined the unphased maps of chromosome X for the karyotypically normal female cell lines in our study (GM12878, IMR90, HMEC,

translocation, der(6)t(6,11)(pter;q) (Figures S7C–S7F). The Hi-C data reveal that the translocation involves the paternal homologs, which cannot be determined with ordinary cytogenetic methods.

We also observed differences in loop structure between homologous autosomes at some imprinted loci. For instance, the *H19/Igf2* locus on chromosome 11 is a well-characterized case

NHEK), the superdomains on X were evident, although the signal was attenuated due to the superposition of signals from active and inactive X chromosomes. When we examined the male HUVEC cell line and the haploid KBM7 cell line, we saw no evidence of superdomains (Figure S7G).

Interestingly, the boundary between the superdomains (ChrX: 115 Mb \pm 500 kb) lies near the macrosatellite repeat *DXZ4* (ChrX: 114,867,433–114,919,088) near the middle of Xq. *DXZ4* is a CpG-rich tandem repeat that is conserved across primates and monkeys and encodes a long noncoding RNA. In males and on the active X, *DXZ4* is heterochromatic, hypermethylated and does not bind CTCF. On the inactive X, *DXZ4* is euchromatic, hypomethylated, and binds CTCF. *DXZ4* has been hypothesized to play a role in reorganizing chromatin during X inactivation (Chadwick, 2008).

There were also significant differences in loop structure between the chromosome X homologs. We observed 27 large “superloops,” each spanning between 7 and 74 Mb, present only on the inactive X chromosome in the diploid map (Figure 7E). The superloops were also seen in all four unphased maps from karyotypically normal XX cells, but were absent in unphased maps from XO and XY cells (Figure S7I). Two of the superloops (chrX:56.8 Mb–*DXZ4* and *DXZ4*–130.9 Mb) were reported previously in a locus-specific study (Horakova et al., 2012).

Like the peak loci of most other loops, nearly all the superloop anchors bind CTCF (23 of 24). The six anchor regions most frequently associated with superloops are large (up to 200 kb). Four of these anchor regions contain whole long noncoding RNA (lncRNA) genes: *loc550643*, *XIST*, *DXZ4*, and *FIRRE*. Three (*loc550643*, *DXZ4*, and *FIRRE*) contain CTCF-binding tandem repeats that only bind CTCF on the inactive homolog.

DISCUSSION

Using the in situ Hi-C protocol, we probed genomic architecture with high resolution; in the case of GM12878 lymphoblastoid cells, better than 1 kb. We observe the presence of contact domains that were too small (median length = 185 kb) to be seen in previous maps. Loci within a domain interact frequently with one another, have similar patterns of chromatin modifications, and exhibit similar long-range contact patterns. Domains tend to be conserved across cell types and between human and mouse. When the pattern of chromatin modifications associated with a domain changes, the domain's long-range contact pattern also changes. Domains exhibit at least six distinct patterns of long-range contacts (subcompartments), which subdivide the two compartments that we previously reported based on low resolution data. The subcompartments are each associated with distinct chromatin patterns. It is possible that the chromatin patterns play a role in bringing about the long-range contact patterns, or vice versa.

Our data also make it possible to create a genome-wide catalog of chromatin loops. We identified loops by looking for pairs of loci that have significantly more contacts with one another than they do with other nearby loci. In our densest map (GM12878), we observe 9,448 loops.

The loops reported here have many interesting properties. Most loops are short (<2 Mb) and strongly conserved across

cell types and between human and mouse. Promoter-enhancer loops are common and associated with gene activation. Loops tend not to overlap; they often demarcate contact domains, and may establish them. CTCF and the cohesin subunits RAD21 and SMC3 associate with loops; each of these proteins is found at over 86% of loop anchors.

The most striking property of loops is that the pair of CTCF motifs present at the loop anchors occurs in a convergent orientation in >90% of cases (versus 25% expected by chance). The importance of motif orientation between loci that are separated by, on average, 360 kb is surprising and must bear on the mechanism by which CTCF and cohesin form loops, which seems likely to involve CTCF dimerization. Experiments in which the presence or orientation of CTCF sites is altered may enable the engineering of loops, domains, and other chromatin structures.

It is interesting to compare our results to those seen in previous reports. The contact domains we observe are similar in size to the “physical domains” that have been reported in Hi-C maps of *Drosophila* (Sexton et al., 2012) and to the “topologically constrained domains” (mean length: 220 kb) whose existence was demonstrated in the 1970s and 1980s in structural studies of human chromatin (Cook and Brazell, 1975; Vogelstein et al., 1980; Zehnbauser and Vogelstein, 1985). On the other hand, the domains we observe are much smaller than the TADs (1 Mb) (Dixon et al., 2012) that have been reported in humans and mice on the basis of lower-resolution contact maps. This is because detecting TADs involves detection of domain boundaries. With higher resolution data, it is possible to detect additional boundaries beyond those seen in previous maps. Interestingly, nearly all the boundaries we observe are associated with either a subcompartment transition (that occur approximately every 300 kb), or a loop (that occur approximately every 200 kb); and many are associated with both.

Our annotation identifies many fewer loops than were reported in several recent high-throughput studies, despite the fact that we have more data. The key reason is that we call peaks only when a pair of loci shows elevated contact frequency relative to the local background—that is, when the peak pixel is enriched as compared to other pixels in its neighborhood. In contrast, prior studies have defined peaks by comparing the contact frequency at a pixel to the genome-wide average (Sanyal et al., 2012; Jin et al., 2013). This latter definition is problematic because many pixels within a domain can be annotated as peaks despite showing no local increase in contact frequency. Papers using the latter definition imply the existence of more than 100,000 loops (1,187 loops were reported in 1% of the genome [Sanyal et al., 2012]) or even more than 1 million loops (reported in a genome-wide Hi-C study [Jin et al., 2013]). The vast majority of the loops annotated by these papers show no enrichment relative to the local background when examined one-by-one and no enrichment with respect to any published Hi-C data set when analyzed using APA (see Extended Experimental Procedures; Figure S8; Data S2). This suggests that these peak annotations may correspond to pairs of loci that lie in the same domain or compartment, but rarely correspond to loops.

We created diploid Hi-C maps by using polymorphisms to assign contacts to distinct chromosomal homologs. We found that the inactive X chromosome is partitioned into two large

superdomains whose boundary lies near the locus of the lncRNA *DXZ4*. We also detect a network of long-range superloops, the strongest of which are anchored at locations containing lncRNA genes (*loc550643*, *XIST*, *DXZ4*, and *FIRRE*). With the exception of *XIST*, all of these lncRNAs contain CTCF-binding tandem repeats that bind CTCF only on the inactive X.

In our original report on Hi-C, we observed that Hi-C maps can be used to study physical models of genome folding, and we proposed a fractal globule model for genome folding at the megabase scale. The kilobase-scale maps reported here allow the physical properties of genome folding to be probed at much higher resolution. We will report such studies elsewhere.

Just as loops bring distant DNA loci into close spatial proximity, we find that they bring disparate aspects of DNA biology—domains, compartments, chromatin marks, and genetic regulation—into close conceptual proximity. As our understanding of the physical connections between DNA loci continues to improve, our understanding of the relationships between these broader phenomena will deepen.

EXPERIMENTAL PROCEDURES

In Situ Hi-C Protocol

All cell lines were cultured following the manufacturer's recommendations. Two to five million cells were crosslinked with 1% formaldehyde for 10 min at room temperature. Nuclei were permeabilized. DNA was digested with 100 units of MboI, and the ends of restriction fragments were labeled using biotinylated nucleotides and ligated in a small volume. After reversal of cross-links, ligated DNA was purified and sheared to a length of ~400 bp, at which point ligation junctions were pulled down with streptavidin beads and prepped for Illumina sequencing. Dilution Hi-C was performed as in Lieberman-Aiden et al. (2009).

3D-FISH

3D DNA-FISH was performed as in Beliveau et al. (2012) with minor modifications.

Hi-C Data Pipeline

All sequence data were produced using Illumina paired-end sequencing. We processed data using a custom pipeline that was optimized for parallel computation on a cluster. The pipeline uses BWA (Li and Durbin, 2010) to map each read end separately to the b37 or mm9 reference genomes; removes duplicate and near-duplicate reads; removes reads that map to the same fragment; and filters the remaining reads based on mapping quality score. Contact matrices were generated at base pair delimited resolutions of 2.5 Mb, 1 Mb, 500 kb, 250 kb, 100 kb, 50 kb, 25 kb, 10 kb, and 5 kb, as well as fragment-delimited resolutions of 500 f, 200 f, 100 f, 50 f, 20 f, 5 f, 2 f, and 1 f. For our largest maps, we also generated a 1 kb contact matrix. Normalized contact matrices are produced at all resolutions using Knight and Ruiz (2012).

Annotation of Domains: Arrowhead

To annotate domains, we apply an "arrowhead" transformation, defined as $A_{i,i+d} = (M_{i,i+d}^* - M_{i,i+d}^e) / (M_{i,i+d}^* + M_{i,i+d}^e)$. M^* denotes the normalized contact matrix (see Figures S2A–S2F). This is equivalent to calculating a matrix equal to $-1 \times (\text{observed/expected} - 1)$, where the expected model controls for local background and distance from the diagonal in the simplest possible way: the "expected" value at $i, i + d$ is simply the mean of the observed values at $i, i - d$ and $i, i + d$. $A_{i,i+d}$ will be strongly positive if locus $i - d$ is inside a domain and locus $i + d$ is not. If the reverse is true, $A_{i,i+d}$ will be strongly negative. If the loci are both inside or both outside a domain, $A_{i,i+d}$ will be close to zero. Consequently, if there is a domain at $[a, b]$, we find that A takes on very negative values inside a triangle whose vertices lie at $[a, a]$, $[a, b]$, and $[(a + b)/2, b]$ and very positive values inside a triangle whose vertices lie at $[(a + b)/2, b]$, $[b, b]$, and $[b, 2b - a]$. The size and positioning of these triangles creates the arrow-

head-shaped feature that replaces each domain in M^* . A "corner score" matrix, indicating each pixel's likelihood of lying at the corner of a domain, is efficiently calculated from the arrowhead matrix using dynamic programming.

Assigning Loci to Subcompartments

To cluster loci based on long-range contact patterns, we constructed a 100 kb resolution interchromosomal contact matrix such that loci from odd chromosomes appeared on the rows, and loci from even chromosomes appeared on the columns. (Intrachromosomal data and data involving chromosome X were excluded.) We cluster this matrix using the Python package *scikit*. For subcompartment B4, the 100 kb interchromosomal matrix for chromosome 19 was constructed and clustered separately, using the same procedure.

Annotation of Peaks: HiCUPS

Our peak-calling algorithm examines each pixel in a Hi-C contact matrix and compares the number of contacts in the pixel to the number of contacts in a series of regions surrounding the pixel. The algorithm thus identifies "enriched pixels" $M_{i,j}^*$ where the contact frequency is higher than expected and where this enrichment is not the result of a larger structural feature. For instance, we rule out the possibility that the enrichment of pixel $M_{i,j}^*$ is the result of L_i and L_j lying in the same domain by comparing the pixel's contact count to an expected model derived by examining the "lower-left" neighborhood. (The "lower-left" neighborhood samples pixels $M_{i',j'}$ where $i \leq i' \leq j' \leq j$; if a pixel is in a domain, these pixels will necessarily be in the same domain.) We require that the pixel being tested contain at least 50% more contacts than expected based on the lower-left neighborhood and the enrichment be statistically significant after correcting for multiple hypothesis testing (False Discovery Rate < 10%). The same criteria are applied to three other neighborhoods. Thus, to be labeled an enriched pixel, a pixel must be significantly enriched relative to four neighborhoods: (1) pixels to its lower-left, (2) pixels to its left and right, (3) pixels above and below, and (4) a donut surrounding the pixel of interest (Figure 3A). The resulting enriched pixels tend to form contiguous interaction regions comprising 5–20 pixels each. We define the "peak pixel" (or simply the "peak") to be the pixel in an interaction region with the most contacts.

Because of the enormous number of pixels that must be examined, this calculation requires weeks of central processing unit (CPU) time to execute. (For instance, at a matrix resolution of 5 kb, the algorithm must be run on 20 billion pixels.) To accelerate it, we created a highly parallelized implementation using general-purpose graphical processing units resulting in a 200-fold speedup.

Aggregate Peak Analysis

We perform APA on 10 kb resolution contact matrices. To measure the aggregate enrichment of a set of putative peaks in a contact matrix, we plot the sum of a series of submatrices derived from that contact matrix. Each of these submatrices is a 210 kb \times 210 kb square centered at a single putative peak in the upper triangle of the contact matrix. The resulting APA plot displays the total number of contacts that lie within the entire putative peak set at the center of the matrix; the entry immediately to the right of center corresponds to the total number of contacts in the pixel set obtained by shifting the peak set 10 kb to the right; the entry two positions above center corresponds to an upward shift of 20 kb and so on. Focal enrichment across the peak set in aggregate manifests as larger values at the center of the APA plot. The APA plots shown only include peaks whose loci are at least 300 kb apart.

ACCESSION NUMBERS

The Gene Expression Omnibus (GEO) accession number for the data sets reported in this paper is GSE63525. The dbGaP accession number for the HeLa data reported in this paper is phs000640.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, eight figures, two data files, and eight tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.021>.

AUTHOR CONTRIBUTIONS

E.L.A. conceived this project. S.S.P.R., M.H.H., E.K.S., and E.L.A. designed experiments. S.S.P.R., E.K.S., I.D.B., A.D.O., and M.H.H. performed Hi-C experiments. E.K.S. and I.D.B. performed 3D-FISH experiments. N.C.D. built the computational pipeline for Hi-C data. N.C.D. and J.T.R. built the visualization system for Hi-C data. S.S.P.R., M.H.H., N.C.D., A.L.S., I.M., E.S.L., and E.L.A. analyzed data. S.S.P.R., M.H.H., N.C.D., E.S.L., and E.L.A. prepared the manuscript.

ACKNOWLEDGMENTS

This paper is dedicated to the memory of Aharon Lieberman. Our work was supported by an NSF Graduate Research Fellowship (DGE0946799 and DGE1144152) to M.H.H., an NIH New Innovator Award (OD008540-01), an NSF Physics Frontier Center (PHY-1427654, Center for Theoretical Biological Physics), an NHGRI CEGS (HG006193), NVIDIA, IBM, Google, a CPRIT Scholar Award (R1304), a McNair Medical Institute Scholar Award, and the President's Early Career Award in Science and Engineering to E.L.A., and an NHGRI grant (HG003067) to E.S.L. We thank Leslie Gaffney, Lauren Solomon, and Bang Wong for assistance with figures; BCM's Integrated Microscopy Core, Michael Mancini, Justin Demmerle, Wendy Salmon, Fabio Stossi, Radhika Dandekar, Sanjay Krishna, and especially Asha Multani for microscopy assistance; Aharon Lieberman, Aviva Presser Aiden, Nicholas Christakis, James Lupski, José Onuchic, Mitchell Guttman, Andreas Gnirke, Louise Williams, Chad Nusbaum, John Bohannon, Olga Dudchenko, and the Aiden laboratory for discussions; and Robbyn Issner and Broad's ENCODE group for several cell lines. The Center for Genome Architecture is grateful to Janice, Robert, and Cary McNair for support. A provisional patent covering in situ Hi-C and related methods has been filed. All sequence data reported in this paper that were not derived from HeLa cells have been deposited at GEO (<http://www.ncbi.nlm.nih.gov/geo/>) (GSE63525). Some of the genome sequences described in this research were derived from a HeLa cell line. Henrietta Lacks, and the HeLa cell line that was established from her tumor cells without her knowledge or consent in 1951, have made significant contributions to scientific progress and advances in human health. We are grateful to Henrietta Lacks, now deceased, and to her surviving family members for their contributions to biomedical research. The HeLa data generated from this research were submitted to the database of Genotypes and Phenotypes (dbGaP) as a substudy under accession number phs000640.

Received: October 12, 2014

Revised: November 5, 2014

Accepted: November 13, 2014

Published: December 11, 2014

REFERENCES

- Ahmadiyeh, N., Pomerantz, M.M., Grisanzio, C., Herman, P., Jia, L., Almendro, V., He, H.H., Brown, M., Liu, X.S., Davis, M., et al. (2010). 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc. Natl. Acad. Sci. USA* **107**, 9742–9746.
- Amano, T., Sagai, T., Tanabe, H., Mizushima, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* **16**, 47–57.
- Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308.
- Beliveau, B.J., Joyce, E.F., Apostolopoulos, N., Yilmaz, F., Fonseka, C.Y., McCole, R.B., Chang, Y., Li, J.B., Senaratne, T.N., Williams, B.R., et al. (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. USA* **109**, 21301–21306.
- Bickmore, W.A. (2013). The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.* **14**, 67–84.
- Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. *Science* **281**, 60–63.
- Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.-L.L., Ruan, Y., Wei, C.-L.L., Ng, H.H., and Liu, E.T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* **18**, 1752–1762.
- Chadwick, B.P. (2008). DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res.* **18**, 1259–1269.
- Cook, P.R., and Brazell, I.A. (1975). Supercoils in human DNA. *J. Cell Sci.* **19**, 261–279.
- Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301.
- Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.-Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19**, 24–32.
- Cullen, K.E., Kladde, M.P., and Seyfred, M.A. (1993). Interaction between transcription regulatory regions of prolactin chromatin. *Science* **261**, 203–206.
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* **295**, 1306–1311.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380.
- Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., et al. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.
- Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R., and Bickmore, W.A. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet.* **4**, e1000039.
- Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **462**, 58–64.
- Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* **7**, 703–713.
- Gavrilov, A.A., Gushchanskaya, E.S., Strelkova, O., Zhironkina, O., Kireev, I.I., Iarovaia, O.V., and Razin, S.V. (2013). Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub. *Nucleic Acids Res.* **41**, 3563–3575.
- Hahn, M.A., Wu, X., Li, A.X., Hahn, T., and Pfeifer, G.P. (2011). Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. *PLoS ONE* **6**, e18844.
- Heidari, N., Phanstiel, D.H., He, C., Grubert, F., Jahanbani, F., Kasowski, M., Zhang, M.Q., and Snyder, M.P. (2014). Genome-wide map of regulatory interactions in the human genome. *Genome Res.* Published online September 16, 2014. <http://dx.doi.org/10.1101/gr.176586.114>.
- Hoffman, M.M., Ernst, J., Wilder, S.P., Kundaje, A., Harris, R.S., Libbrecht, M., Giardine, B., Ellenbogen, P.M., Bilmes, J.A., Birney, E., et al. (2013). Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res.* **41**, 827–841.
- Horakova, A.H., Moseley, S.C., McLaughlin, C.R., Tremblay, D.C., and Chadwick, B.P. (2012). The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. *Hum. Mol. Genet.* **21**, 4367–4377.
- Hou, C., Zhao, H., Tanimoto, K., and Dean, A. (2008). CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc. Natl. Acad. Sci. USA* **105**, 20398–20403.
- Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294.

- Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30**, 90–98.
- Knight, P., and Ruiz, D. (2012). A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* Published online October 26, 2012. <http://dx.doi.org/10.1093/imanum/drs019>.
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595.
- Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.
- Mukherjee, S., Erickson, H., and Bastia, D. (1988). Enhancer-origin interaction in plasmid R6K involves a DNA loop mediated by initiator protein. *Cell* **52**, 375–383.
- Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385.
- 1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65.
- Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. *Cell* **137**, 1194–1211.
- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113.
- Schleif, R. (1992). DNA looping. *Annu. Rev. Biochem.* **61**, 199–223.
- Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G.D., Marshall, A., Flícek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335–348.
- Sexton, T., Schober, H., Fraser, P., and Gasser, S.M. (2007). Gene regulation through nuclear organization. *Nat. Struct. Mol. Biol.* **14**, 1049–1055.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472.
- Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* **20**, 2349–2354.
- Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10**, 1453–1465.
- Vogel, M.J., Guelen, L., de Wit, E., Peric-Hupkes, D., Lodén, M., Talhout, W., Feenstra, M., Abbas, B., Classen, A.K., and van Steensel, B. (2006). Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res.* **16**, 1493–1504.
- Vogelstein, B., Pardoll, D.M., and Coffey, D.S. (1980). Supercoiled loops and eucaryotic DNA replication. *Cell* **22**, 79–85.
- Xie, X., Mikkelsen, T.S., Gnirke, A., Lindblad-Toh, K., Kellis, M., and Lander, E.S. (2007). Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc. Natl. Acad. Sci. USA* **104**, 7145–7150.
- Yaffe, E., and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet.* **43**, 1059–1065.
- Zehnbaumer, B.A., and Vogelstein, B. (1985). Supercoiled loops and the organization of replication and transcription in eukaryotes. *BioEssays* **2**, 52–54.

X Chromosome Reactivation Dynamics Reveal Stages of Reprogramming to Pluripotency

Vincent Pasque,^{1,6} Jason Tchieu,^{2,6} Rahul Karnik,³ Molly Uyeda,¹ Anupama Sadhu Dimashkie,¹ Dana Case,¹ Bernadett Papp,¹ Giancarlo Bonora,¹ Sanjeet Patel,¹ Ritchie Ho,¹ Ryan Schmidt,¹ Robin McKee,¹ Takashi Sado,⁴ Takashi Tada,⁵ Alexander Meissner,³ and Kathrin Plath^{1,*}

¹Department of Biological Chemistry, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, Jonsson Comprehensive Cancer Center, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA

²Developmental Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

³Department of Stem Cell and Regenerative Biology, Harvard University, Harvard Stem Cell Institute, Broad Institute of MIT and Harvard, Cambridge, MA 02138, USA

⁴Department of Advanced Bioscience, Graduate School of Agriculture, Kinki University, 3327-204 Nakamachi, Nara, 631-8505, Japan

⁵Department of Stem Cell Engineering, Stem Cell Research Center, Institute for Frontier Medical Sciences, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan

⁶Co-first author

*Correspondence: kplath@mednet.ucla.edu

<http://dx.doi.org/10.1016/j.cell.2014.11.040>

SUMMARY

Reprogramming to iPSCs resets the epigenome of somatic cells, including the reversal of X chromosome inactivation. We sought to gain insight into the steps underlying the reprogramming process by examining the means by which reprogramming leads to X chromosome reactivation (XCR). Analyzing single cells in situ, we found that hallmarks of the inactive X (Xi) change sequentially, providing a direct readout of reprogramming progression. Several epigenetic changes on the Xi occur in the inverse order of developmental X inactivation, whereas others are uncoupled from this sequence. Among the latter, DNA methylation has an extraordinary long persistence on the Xi during reprogramming, and, like *Xist* expression, is erased only after pluripotency genes are activated. Mechanistically, XCR requires both DNA demethylation and *Xist* silencing, ensuring that only cells undergoing faithful reprogramming initiate XCR. Our study defines the epigenetic state of multiple sequential reprogramming intermediates and establishes a paradigm for studying cell fate transitions during reprogramming.

INTRODUCTION

Understanding the mechanisms by which the identity of a cell is established and maintained is a key goal of contemporary biology. Somatic cells can be reprogrammed into induced pluripotent stem cells (iPSCs) through transcription factor expression (Takahashi and Yamanaka, 2006). This process entails profound changes in genome organization, histone modifications, DNA methylation, and gene expression (reviewed in

Apostolou and Hochedlinger, 2013). Questions of outstanding interest are whether reprogramming proceeds through specific stages that can be defined based on epigenetic features and how and in what order the epigenetic features gradually acquired during differentiation are reversed during reprogramming. One approach to address these questions is to focus on events for which the sequence of epigenetic changes that occur during differentiation is well defined and to ask how it is reversed during reprogramming to iPSCs.

X chromosome inactivation (XCI) is induced upon differentiation of female mouse pluripotent cells and leads to the inactivation of one of the two X chromosomes (reviewed in Lee and Bartolomei, 2013; Barakat and Gribnau, 2010; Chow and Heard, 2009). The sequence of epigenetic events accompanying the silencing of the X chromosome during differentiation has been examined extensively (Chow and Heard, 2009). These events include an initiation phase characterized by the coating of the future inactive X chromosome (Xi) by the large noncoding RNA *Xist*, which creates a nuclear compartment devoid of RNA polymerase II (Chaumeil et al., 2006) and leads to transient recruitment of the Polycomb Repressive Complex 2 (PRC2) and the deposition of the repressive histone mark H3K27me3 by its catalytic subunit EZH2 (Plath et al., 2003; Silva et al., 2003), closely followed by gene silencing. Later in differentiation, these events are followed by incorporation of the repressive histone variant macroH2A1 and DNA methylation, stabilizing the silenced state (Gendrel et al., 2012; Mermoud et al., 1999). Thus, once established, the Xi is extraordinarily stable and is only reversed in a process termed X chromosome reactivation (XCR), which, in embryos, is limited to the inner cell mass and to germ cells (Lee and Bartolomei, 2013). XCR results in erasure of Xi-heterochromatin marks, and, importantly, can also be induced experimentally by reprogramming of female mouse somatic cells to iPSCs and somatic cell nuclear transfer (SCNT) (Maherali et al., 2007; Eggan et al., 2000). It is known that XCR is a late event during reprogramming to iPSCs (Payer et al., 2013;

Stadtfield et al., 2008), but the exact dynamics of XCR and how the epigenetic hallmarks of the Xi change in this process have remained unclear.

Most insight into the molecular events of reprogramming to iPSCs have been gained from gene expression studies of populations of cells undergoing reprogramming and of subpopulations isolated using cell surface markers (O'Malley et al., 2013; Golipour et al., 2012; Polo et al., 2012; Samavarchi-Tehrani et al., 2010; Stadtfield et al., 2008; Mikkelsen et al., 2008; reviewed in Buganim et al., 2013). These studies indicated that reprogramming is a multistep process with two predominant "waves" of gene expression changes: an early wave marked by enhanced proliferation and a mesenchymal-to-epithelial transition (MET), characterized by *Cdh1* (E-cadherin) expression (Polo et al., 2012; Samavarchi-Tehrani et al., 2010; Li et al., 2010), and a late wave, characterized by reactivation of pluripotency genes such as *Nanog* (O'Malley et al., 2013; Buganim et al., 2012; Golipour et al., 2012; Polo et al., 2012). The variable latency and relatively low efficiency by which individual cells reprogram have also encouraged gene expression measurements at the single-cell level at various stages of reprogramming and in clonal late intermediates. These experiments have argued for a sequence of stochastic transcriptional changes early in reprogramming, where expression programs vary dramatically between individual cells, eventually leading to hierarchical activation of pluripotency genes during the final phase, which, however, may occur through multiple paths (Buganim et al., 2012; Polo et al., 2012; Parchem et al., 2014).

Despite these advances, further molecular insight into the reprogramming path and a continuous view of the molecular events and stages leading to pluripotency would benefit from alternative approaches. In situ temporal analyses that integrate the position of cells within their native reprogramming environment, as well as the level of proteins and chromatin marks and their subcellular localization, may be particularly useful. Given that reprogramming to iPSCs is associated with XCR, and in light of the detailed characterization of sequential steps of XCI during differentiation, the reprogramming process provides an unprecedented opportunity to study XCR. In turn, the Xi provides an exceptional possibility to characterize the dynamics of the reversal of epigenetic marks during reprogramming.

Here, we followed epigenetic changes on the Xi during reprogramming to iPSCs in individual cells using detailed, high-resolution in situ time course analyses to address the question of whether XCR and somatic cell reprogramming follow a precise sequence of epigenetic changes. Due to the sheer size of the X chromosome, this analysis can be done at the single-cell level using immunofluorescence and RNA FISH approaches, allowing for the identification of reprogramming stages that have been elusive in transcriptional and chromatin studies to date. Our work demonstrates that the epigenetic state of the Xi changes sequentially throughout reprogramming, along with global changes in chromatin character. To shed light on the mechanisms by which XCR takes place, we used genetically manipulated somatic cells and examined the role played by *Cdh1*, *Nanog*, *Xist*, *Tsix*, *Tet1*, *Tet2*, and DNA methylation. The highly reproducible sequence of epigenetic steps leading to XCR and induced pluripotency provides a simple readout of reprogram-

ming progression and a basis for studying cell fate transitions during reprogramming.

RESULTS

Reprogramming Steps Defined by the Dynamics of Xi Chromatin Marks

To define epigenetic steps of XCR and reprogramming, we determined the dynamics of Xi hallmarks during the establishment of pluripotency in mouse embryonic fibroblasts (MEFs). We induced female MEFs to reprogram with retroviruses encoding *Oct4*, *Sox2*, and *Klf4* and analyzed single cells in their native reprogramming environment throughout detailed time courses every other day for 1–3 weeks using multicolor immunostaining (Figure 1A). This allowed us to assess the state of the Xi and of evolving global epigenetic states in any cell of the reprogramming cultures and to delineate the sequence of epigenetic events during reprogramming relative to other markers.

We first analyzed the dynamics of PRC2 on the Xi. EZH2 did not accumulate on the Xi within the first 6 days of reprogramming (Figures 1B and 1C, i). However, after CDH1 (E-cadherin) became expressed, which marks the MET (Li et al., 2010), and before the pluripotency factor NANOG was detectable, a strong nuclear EZH2 staining focus characteristic of Xi accumulation ($\text{Xi}^{\text{EZH2}+}$) arose in a small fraction of the cells (Figures 1B, 1C, ii, and 1D). The same result was obtained for SUZ12, another PRC2 subunit, and the PRC2-recruitment factor JARID2 (da Rocha et al., 2014) (Figure S1A available online). The $\text{Xi}^{\text{EZH2}+}$ was restricted to CDH1+ cells (Figure 1E) and only occurred in a subset of CDH1+ cells, around 50% at day 10 of reprogramming. These findings show that $\text{Xi}^{\text{EZH2}+}$ arises after an epithelial cell character is established during reprogramming, indicative of the existence of a reprogramming stage immediately downstream to MET that is more restrictive than CDH1 expression. In agreement with this, $\text{Xi}^{\text{EZH2}+}$ was also present in known late reprogramming intermediates such as pre-iPSCs (Figures S1B–S1D) and was only detectable in reprogramming cultures cotransduced with viruses encoding *Oct4*, *Sox2*, and *Klf4* with or without *cMyc*, but not when fewer reprogramming factors were employed (Figure S1E), demonstrating that the PRC2 composition of the Xi only changes when reprogramming factor combinations able to induce pluripotency are used.

Notably, we also observed that the level of nuclear EZH2 (i.e., on autosomes) gradually increased during reprogramming in both male and female cells, which was initiated specifically in a subset of CDH1+ cells before $\text{Xi}^{\text{EZH2}+}$ was induced (Figure 1C, progression from i to iv). However, ectopic EZH2 expression in female MEFs did not induce $\text{Xi}^{\text{EZH2}+}$ (Figure S1F), indicating that the global EZH2 increase during reprogramming is not sufficient for $\text{Xi}^{\text{EZH2}+}$. These results reveal that, downstream of MET, PRC2 is gradually upregulated at the global level, irrespective of sex chromosome content, and additionally relocates to the Xi in female cells, providing a direct readout of reprogramming progression.

To uncover whether CDH1-positive (CDH1+)/ $\text{Xi}^{\text{EZH2}+}$ cells are intermediates on the path to the NANOG-positive (NANOG+) reprogramming stage, we determined the presence of $\text{Xi}^{\text{EZH2}+}$ in the first cells that express the NANOG protein during

reprogramming. We found that NANOG activation initiated within a subset of $\text{Xi}^{\text{EZH2}+}$ colonies, with nearly all NANOG+ cells that first appeared in reprogramming cultures carrying the $\text{Xi}^{\text{EZH2}+}$ (Figures 1C, iii, and 1F). Later in reprogramming, and usually in large NANOG+ colonies, almost all NANOG+ cells lacked $\text{Xi}^{\text{EZH2}+}$ (Figures 1C, iv, and 1F), and the absolute number of $\text{Xi}^{\text{EZH2}+}$ cells and colonies decreased accordingly (data not shown). Moreover, NANOG+ cells were initially surrounded by NANOG-negative ($\text{NANOG-}/\text{Xi}^{\text{EZH2}+}$) cells (Figure S1G), which is consistent with the induction of NANOG occurring in a subset of $\text{CDH1}+/\text{Xi}^{\text{EZH2}+}$ cells followed by removal of $\text{Xi}^{\text{EZH2}+}$ within NANOG+ colonies.

H3K27me3, the downstream mark of PRC2, enriched on the Xi in NANOG- cells, was lost from the Xi exclusively within NANOG+ cells with kinetics slightly delayed compared to the loss of $\text{Xi}^{\text{EZH2}+}$, such that NANOG+ cells with $\text{Xi}^{\text{H3K27me3}+}$ but without $\text{Xi}^{\text{EZH2}+}$ could be briefly detected (Figures 1G and 1H). These data suggest that the loss of $\text{Xi}^{\text{H3K27me3}+}$ is a consequence of the removal of EZH2 from the Xi. Thus, NANOG expression precedes both loss of $\text{Xi}^{\text{EZH2}+}$ and $\text{Xi}^{\text{H3K27me3}+}$.

Taken together, our findings suggest that cells go through defined epigenetic steps as they progress toward pluripotency. Specifically, we reveal four steps by simply following PRC2, at the Xi-specific and global level, relative to CDH1 and NANOG. Downstream of MET, PRC2 proteins increase in overall levels and accumulate on the Xi in a subset of $\text{CDH1}+$ cells. Then, a subset of $\text{CDH1}+/\text{Xi}^{\text{EZH2}+}$ cells reactivates NANOG, which precedes EZH2 and H3K27me3 removal from the Xi specifically in these NANOG+ cells (Figure 1I). Importantly, the reacquisition of $\text{Xi}^{\text{EZH2}+}$ represents the inversed sequence of events of developmental XCI, where PRC2 accumulates on the Xi immediately after *Xist* RNA initially coats the X and disappears from the Xi later in differentiation (Plath et al., 2003; Silva et al., 2003), suggesting that $\text{Xi}^{\text{EZH2}+}$ reflects the extent of reversal of the differentiated state established during reprogramming.

The histone variant macroH2A1 associates with the Xi late during developmental XCI and has been shown to act as a barrier to reprogramming (Pasque et al., 2012). We found that the $\text{Xi}^{\text{macroH2A1}+}$ of MEFs was maintained during reprogramming until $\text{Xi}^{\text{EZH2}+}$ was lost (Figures S1H–S1K). Unexpectedly, the global level of macroH2A1 first increased from the somatic level before dropping again to the lower level of pluripotent cells in both female and male cells (Figure S1I; data not shown), indicating that an epigenetic mark associated with resistance to reprogramming is transiently induced during the reprogramming process (Figure S1O). Altogether, these results strengthen the conclusion that the Xi-specific and global epigenetic states of cells define multiple stages of reprogramming. Unlike $\text{Xi}^{\text{EZH2}+}$, the kinetics of $\text{Xi}^{\text{macroH2A1}+}$ loss during reprogramming does not represent the reversed sequence of developmental $\text{Xi}^{\text{macroH2A1}+}$ dynamics, suggesting that distinct mechanisms regulate the temporal Xi accumulation of different epigenetic marks during reprogramming.

Subpopulations with Increased Reprogramming Capacity Recapitulate Xi Events

To test whether the steps identified based on fixed cultures represent dynamics of cells that would, if not fixed, continue

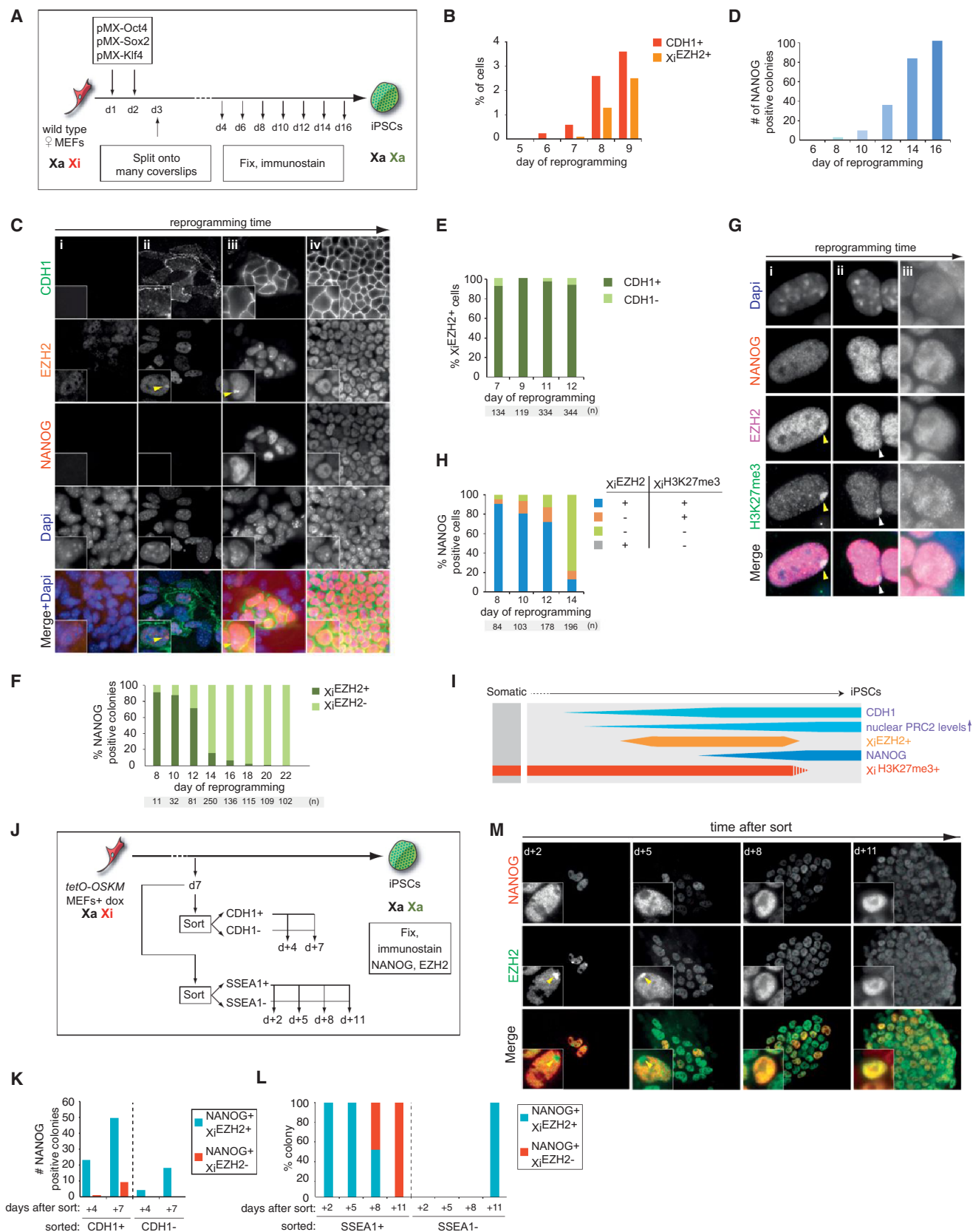
along the path to pluripotency, we considered the use of pluripotency reporters such as *Oct4-GFP* or *Nanog-GFP*. However, we found that their activation occurred well after the endogenous NANOG protein was detectable, at a time when EZH2 is already removed from the Xi (Figures S1L–S1N), precluding their use for monitoring reprogramming events that occur when NANOG becomes initially expressed.

Instead, we asked whether $\text{NANOG}+/\text{Xi}^{\text{EZH2}+}$ cells arise from $\text{CDH1}+$ cells by sorting $\text{CDH1}+$ and CDH1- cells at day 7 of reprogramming and assessing their ability to give rise to $\text{NANOG}+/\text{Xi}^{\text{EZH2}+}$ cells after replating an equal number of both sorted cell populations (Figures 1J and S1P). We found that $\text{CDH1}+$ -sorted cells preferentially gave rise to $\text{NANOG}+/\text{Xi}^{\text{EZH2}+}$ colonies compared to $\text{NANOG}+/\text{Xi}^{\text{EZH2}-}$ colonies (Figure 1K), supporting the conclusion that $\text{NANOG}+/\text{Xi}^{\text{EZH2}+}$ cells originate from $\text{CDH1}+$ cells. Furthermore, in the time frame considered, replated CDH1- -sorted cells also proceeded to the $\text{NANOG}+/\text{Xi}^{\text{EZH2}+}$ state but with delayed kinetics and reduced efficiency (Figure 1K), which is in agreement with the notion that cells reprogram with variable latencies (Hanna et al., 2009). We also performed sorting experiments employing SSEA1, a marker of a reprogramming intermediate arising within $\text{CDH1}+$ cells (Polo et al., 2012) (Figures 1J and S1Q). As expected, shortly after replating, $\text{NANOG}+$ colonies were detected specifically from the $\text{SSEA1}+$ population, and cells within these colonies were initially exclusively $\text{Xi}^{\text{EZH2}+}$ (Figures 1L and 1M). Remarkably, as these $\text{NANOG}+$ colonies grew bigger over time, they completely lost $\text{Xi}^{\text{EZH2}+}$ (Figures 1L and 1M). SSEA1- cells gave rise to $\text{NANOG}+$ cells later, and these were all first $\text{Xi}^{\text{EZH2}+}$ (Figure 1L). Therefore, we conclude that the reprogramming steps defined based on our fixed time courses correctly capture the trajectory of cells moving toward pluripotency.

XCR Occurs after Loss of *Xist* RNA in $\text{NANOG}+$ Cells

To determine the dynamics of *Xist* RNA, the key regulator of developmental XCI, during reprogramming, we combined immunostainings with RNA FISH for *Xist*. Early in reprogramming, virtually all cells showed *Xist* RNA coating, detectable as a large “cloud” of RNA FISH signal (Figure 2A, i). At late reprogramming time points, *Xist* RNA was specifically absent from the Xi within $\text{NANOG}+$ colonies, whereas NANOG- cells still exhibited *Xist* RNA coating (Figure 2A, iii). However, the first $\text{NANOG}+$ cells to appear in culture were always $\text{Xi}^{\text{Xist}+}$ (Figures 2A, ii, and 2B), indicating that *Xist* repression follows NANOG activation. Furthermore, we found that $\text{Xi}^{\text{EZH2}+}$ in $\text{NANOG}+$ cells highly correlated with the presence of *Xist* RNA and that their loss occurred with similar dynamics (Figures S2A and S2B), which is consistent with *Xist*-dependent recruitment of PRC2 to the Xi (Plath et al., 2003).

Next, we used RNA FISH to examine when genes on the Xi reactivate during reprogramming (Figure S2C) and found that cells mostly displayed monoallelic expression of the X-linked genes *Mecp2*, *Atrx*, *Gpc4*, and *Rlim* when $\text{NANOG}+$ cells first appeared, indicative of maintenance of XCI in these cells (Figures 2C and S2D–S2F). Later in reprogramming, $\text{NANOG}+$ cells exhibited biallelic expression of these genes, a sign of XCR (Figures 2C and S2D–S2F). For all tested genes, reactivation occurred with delay relative to the loss of *Xist* RNA. We conclude that XCR is a very late event of reprogramming that occurs in a



(legend on next page)

coordinated fashion along the chromosome after *Xist* RNA coating has disappeared. Our results also suggest that XCR takes place independently in multiple cells of a given NANOG+ colony because all cells in NANOG+ colonies initially are Xi^{Xist+}/Xi^{EZH2+} , whereas, at later reprogramming time points, the cells in larger NANOG+ colonies are not (Figure S2A).

To confirm that XCR occurs only in NANOG+ cells, we performed two additional assays. First, we observed that the exclusion of RNA polymerase II from the Xi domain was maintained in all NANOG+ or NANOG− cells that carried $Xi^{H3K27me3+}$ (Figure 2D), which is consistent with the maintenance of silencing at these reprogramming stages. Second, we examined the expression of an Xi-linked GFP reporter (Maherali et al., 2007) and did not detect GFP reporter reactivation in NANOG− cells, whereas NANOG+ cells consistently expressed GFP at late reprogramming time points (Figure S2G). We conclude that, late in reprogramming, NANOG expression precedes the loss of *Xist* RNA, which coincides with loss of Xi^{EZH2+} and occurs before XCR (Figure 2E).

Reprogramming Reverses the Developmental Sequence of *Tsix* Expression

To establish the dynamics of activation of *Tsix* (transcribed antisense to *Xist*) RNA, a critical regulator of *Xist* during initiation of XCI (Lee and Bartolomei, 2013), we used strand-specific RNA FISH and found that *Tsix* was not expressed during the early stages of reprogramming (i.e., in NANOG− cells) or in the first NANOG+ cells that appear (Figure 2F). Within maturing NANOG+ cells, however, *Tsix* became first monoallelically expressed in cells still carrying Xi^{Xist+} . The monoallelic *Tsix* signal occurred

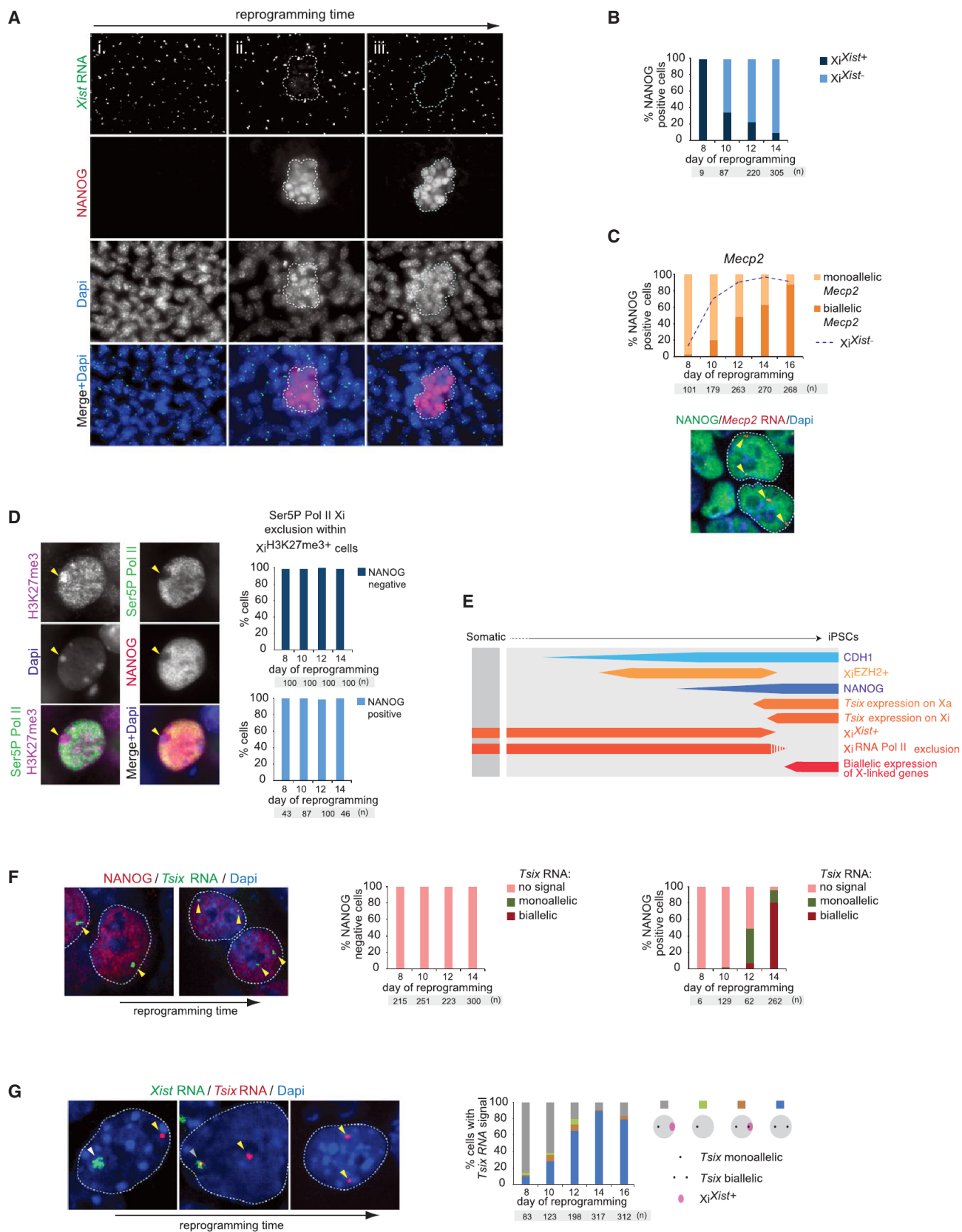
specifically from the active X chromosome (Xa), as it never overlapped with Xi^{Xist+} (Figure 2G). *Tsix* activation on the Xi took place later, at the very tail end of Xi^{Xist} loss (Figures 2G and S2H). Together, these results show that reprogramming to pluripotency recapitulates the expression of *Tsix* in the reverse order from that of developmental XCI, where *Tsix* is first downregulated on the future Xi and then becomes repressed on the Xa (Lee and Lu, 1999) (Figure 2E).

Kinetics of XCR in Relation to Pluripotency Gene Activation

Given the stepwise changes of Xi hallmarks late in reprogramming, we aimed to determine the dynamics of these features in relation to the activation of pluripotency-associated factors. In agreement with the reported hierarchical activation of pluripotency factors late in reprogramming based on single-cell transcript analysis (Buganim et al., 2012), we observed the sequential induction of the pluripotency factors ESRRB, REX1, DPPA4, and PECAM1 at the single-cell level using multicolor immunostaining, which only occurred in NANOG+ cells (Figures S3A–S3F). In addition, silencing of the reprogramming factor-expressing retroviruses can be placed early in this hierarchy at around the time of ESRRB/REX1 activation (Figures S3G–S3I), which is consistent with the shift to endogenous pluripotency factor activation. Xi^{EZH2+} was lost after REX1 expression and just after DPPA4 activation (Figures 3A and 3B). PECAM1 expression, which is very late in the pluripotency factor hierarchy, marked cells that are devoid of Xi^{Xist+} and Xi^{EZH2+} (Figures 3C and 3D). Consistent with a delay of XCR relative to *Xist* RNA loss, XCR took place after DPPA4 activation, as small DPPA4+ colonies

Figure 1. Time Course Analysis of Xi^{EZH2+} during Reprogramming to Pluripotency

- (A) Diagram of reprogramming time course experiments. In all experiments, results for female cells are displayed except when stated otherwise, and the time points and number (n) of cells or colonies counted are given in each subfigure.
- (B) Quantitation of the proportion of CDH1+ or Xi^{EZH2+} cells at indicated reprogramming time points. 100 cells in three randomly chosen microscopic fields were counted per time point.
- (C) Multicolor immunostaining for CDH1 (green in merge), EZH2 (orange), and NANOG (red) at different stages of reprogramming. Dapi staining (blue) marks nuclei. Unlike MEFs (i), cells with elevated nuclear levels of EZH2 and Xi^{EZH2+} (arrowhead) are seen within CDH1+ cells during reprogramming starting around day 7 of reprogramming (ii). (iii) NANOG+ colonies are first marked by Xi^{EZH2+} and elevated EZH2 levels in the nucleus (image from day 9). (iv) Later, NANOG+ colonies become larger and are characterized by high nuclear EZH2 levels without Xi^{EZH2+} (image from day 14).
- (D) Number of NANOG+ colonies throughout reprogramming (a colony is defined as four or more closely localized cells).
- (E) Proportion of Xi^{EZH2+} cells with and without CDH1 expression during reprogramming.
- (F) Proportion of NANOG+ colonies with or without Xi^{EZH2+} at indicated time points. All NANOG+ colonies present in the reprogramming cultures were counted up to day 14 and only a subset thereafter.
- (G) Multicolor immunostaining for EZH2 (magenta in merge) and NANOG (red) in combination with H3K27me3 (green). The images depict various states of Xi^{EZH2+} and $Xi^{H3K27me3+}$ in NANOG+ cells quantified in (H). During reprogramming, (i) NANOG+/ Xi^{EZH2+} cells are initially $Xi^{H3K27me3+}$ (ii) and, at a later time point, become $Xi^{EZH2-}/Xi^{H3K27me3+}$ for a very short time and subsequently become (iii) $Xi^{EZH2-}/Xi^{H3K27me3-}$. Yellow and white arrowheads indicate $Xi^{EZH2+}/Xi^{H3K27me3+}$ and $Xi^{EZH2-}/Xi^{H3K27me3+}$ patterns, respectively.
- (H) Quantitation of the immunostaining experiment in (G), giving the proportion of NANOG+ cells with Xi^{EZH2+} or $Xi^{H3K27me3+}$ at indicated time points.
- (I) Summary of Xi and global dynamics of PRC2 and H3K27me3 during reprogramming, relative to CHD1 and NANOG expression. Female-specific features are shown in orange/red, and those occurring in both female and male reprogramming are shown in blue. The width of the boxes represents the level of the epigenetic mark considered.
- (J) Experimental design for the isolation and characterization of CDH1+/- or SSEA1+/- reprogramming subpopulations.
- (K) Number of NANOG+ colonies with or without Xi^{EZH2+} in CDH1+ and CDH1− sorted cell populations isolated as shown in (J), at indicated days after replating.
- (L) Proportion of NANOG+ colonies with or without Xi^{EZH2+} in SSEA1+ and SSEA1− sorted cell populations isolated as shown in (J), at indicated days after replating. n = 6 for each SSEA1+ time point, and n = 1 for the SSEA1− count at +d11. Colonies appearing in SSEA1+ replated cells become larger throughout this time course as shown in (M).
- (M) Visualization of Xi^{EZH2+} changes in replated SSEA1+ reprogramming intermediates over time from the experiment shown in (J) and (L). Replated cells were immunostained for EZH2 (green in merge) and NANOG (red) at the indicated days. Note the increase in colony size and disappearance of Xi^{EZH2+} (yellow arrowhead) with time in culture.
- See also Figure S1.



(legend on next page)

expressed the X-linked gene *Atrx* monoallelically and large colonies expressed the X-linked gene mostly biallelically (Figure 3E). Together, these results demonstrate the molecular timeline of Xi changes relative to hierarchical pluripotency gene activation (summarized in Figure 3F).

Sequential Xi States Are Conserved across Different Reprogramming Systems

To establish whether the sequential changes on the Xi during reprogramming are specific to the reprogramming system used, we reprogrammed MEFs carrying a single dox-inducible, polycistronic reprogramming cassette encoding *Oct4*, *Sox2*, *Klf4*, and *cMyc* in a defined locus instead of retroviral infection. In addition, we reprogrammed another starting cell type, mouse embryonic endoderm cells, and also used different culture conditions. The dynamics of $\text{Xi}^{\text{EZH2}+}$ and sequential pluripotency gene activation were reproduced in each case (Figures S3J–S3N). We conclude that the epigenetic states identified represent fundamental changes inherent to reprogramming, applicable to multiple starting cell types and reprogramming systems.

CDH1 and NANOG Are Required, but Not Sufficient, for the Efficient Induction of Reprogramming Steps Leading to XCR

Our data indicate that CDH1 expression marks the cells that subsequently induce $\text{Xi}^{\text{EZH2}+}$ and NANOG and that only those cells that activate NANOG are fated to induce *Xist* loss, *Tsix* activation on the Xi and Xa, pluripotency-associated factor activation, and XCR, suggesting that both CDH1 and NANOG are critical for this hierarchy of events. To address the role of CDH1 and NANOG for epigenetic changes taking place downstream of their expression, we performed both knockdown and overexpression experiments. We found that knockdown of *Cdh1* during reprogramming with shRNAs decreased the number of $\text{Xi}^{\text{EZH2}+}$, NANOG+, and DPPA4+ colonies (Figures 4A and S4A). In contrast, *Cdh1* overexpression did not promote any of the epigenetic events that normally take place downstream of CDH1 induction (Figures S4B–S4E). Depletion of *Nanog* transcripts during reprogramming using an inducible shRNA (Figures S4F and S4G) did not prevent CDH1 activation, global upregulation of EZH2, and $\text{Xi}^{\text{EZH2}+}$ (Figures S4H and S4I), in agreement with

its induction later during reprogramming (Silva et al., 2009). By contrast, the activation of *Tsix* on the Xa and Xi, as well as of pluripotency-associated transcription factors, was strongly reduced by *Nanog* depletion (Figures S4J–S4M). The lack of biallelic *Tsix* expression in the absence of *Nanog* also suggests that XCR was impaired without *Nanog*. Thus, NANOG orchestrates the efficient transition through the later molecular events, including XCR, although the requirement for *Nanog* can be bypassed (Carter et al., 2014; Schwarz et al., 2014). Overexpression of *Nanog* late in reprogramming promoted steps toward XCR as judged by the increased number of DPPA4+/ $\text{Xi}^{\text{EZH2}+}$ colonies, but not those before NANOG is normally induced (i.e., $\text{Xi}^{\text{EZH2}+}$) (Figures 4B, S4N, and S4O). However, most NANOG-overexpressing cells did not induce the subsequent reprogramming steps. Together, these results demonstrate that both *Cdh1* and *Nanog* are required, but not sufficient, for the induction of the epigenetic events leading to XCR.

The above finding raised the question of whether XCR represents a barrier to reprogramming. To test this, we obtained a large number of female and male MEFs preparations from four independent litters and measured the efficiency with which NANOG+, DPPA4+, or PECAM1+ colonies formed, without prior knowledge of the sex. This experiment revealed no difference in the reprogramming efficiency between male and female MEFs in KSR or FBS culture media and in the transition to different reprogramming stages (Figures 4C and 4D). Thus, even though the Xi represents the most extreme form of facultative heterochromatin, XCR does not limit reprogramming to induced pluripotency.

Requirement of *Xist* Silencing, but Not *Tsix* Expression, for XCR

To examine the molecular mechanism of XCR during reprogramming, we focused on the requirement for *Xist* and for *Tsix*, its negative regulator in pluripotent cells (Lee and Bartolomei, 2013). Despite *Tsix* becoming expressed on the Xi as *Xist* RNA disappears (Figure 2), deletion of *Tsix* did not alter the kinetics of *Xist* repression in NANOG+ cells (Figures 5A and 5B), indicating that *Tsix* does not negatively regulate *Xist* at the end of reprogramming. Conversely, to test whether repression of *Xist* RNA is required for XCR, we ectopically expressed *Xist* from the Xi during reprogramming (Figure 5C). Constitutive *Xist*

Figure 2. Kinetics of *Xist* and *Tsix* RNA and XCR during Reprogramming

- (A) Representative images of *Xist* RNA (green in merge), NANOG (red), and Dapi (blue) from immunoFISH stainings at different time points of reprogramming reflecting different states of $\text{Xi}^{\text{Xist}+}$ and NANOG expression as determined in (B). Dotted lines indicate the position of NANOG+ colonies across different channels. (i) day 8, (ii) day 10, and (iii) day 14. Each image represents a series of ten Z-sections merged onto a single plane.
- (B) Proportion of NANOG+ cells with or without $\text{Xi}^{\text{Xist}+}$ at different time points.
- (C) ImmunoFISH analysis of NANOG expression and nascent transcripts of the X-linked gene *Mecp2* (seen as a strong pinpoint) during reprogramming. In the image, the biallelic *Mecp2* expression pattern is indicated (two arrowheads), and the proportion of NANOG+ cells with mono- or biallelic *Mecp2* expression is given below. The dotted line indicates the proportion of $\text{Xi}^{\text{Xist}-}$ cells from the same time course.
- (D) Images depict an immunostaining for H3K27me3, Ser5P polymerase II (Ser5P Pol II), and NANOG at day 12 of reprogramming. Quantification gives the proportion of Ser5P Pol II Xi-exclusion cells in NANOG– (top) or NANOG+ (bottom) cells that also display $\text{Xi}^{\text{H3K27me3}+}$ at indicated time points.
- (E) Summary of reprogramming stages related to this figure, displayed as described in Figure 1. Dashed lines indicate the window of time we narrowed down for the feature to occur or disappear.
- (F) ImmunoFISH analysis as in (C), except for NANOG (red) and *Tsix* RNA (green). Proportion of NANOG– (left) and NANOG+ (right) cells, respectively, displaying monoallelic, biallelic, or no *Tsix* RNA FISH signal.
- (G) RNA FISH analysis of the relationship between *Xist* and *Tsix* RNA in reprogramming. In the images, yellow arrowheads represent *Tsix* expression without *Xist* RNA present on the same X chromosome, and the white and gray arrowheads represent an *Xist* RNA cloud that does not or does overlap a *Tsix* signal, respectively. The quantification of cells with *Tsix* expression showing mono- or biallelic *Tsix* expression with and without $\text{Xi}^{\text{Xist}+}$ is shown. See also Figure S2.

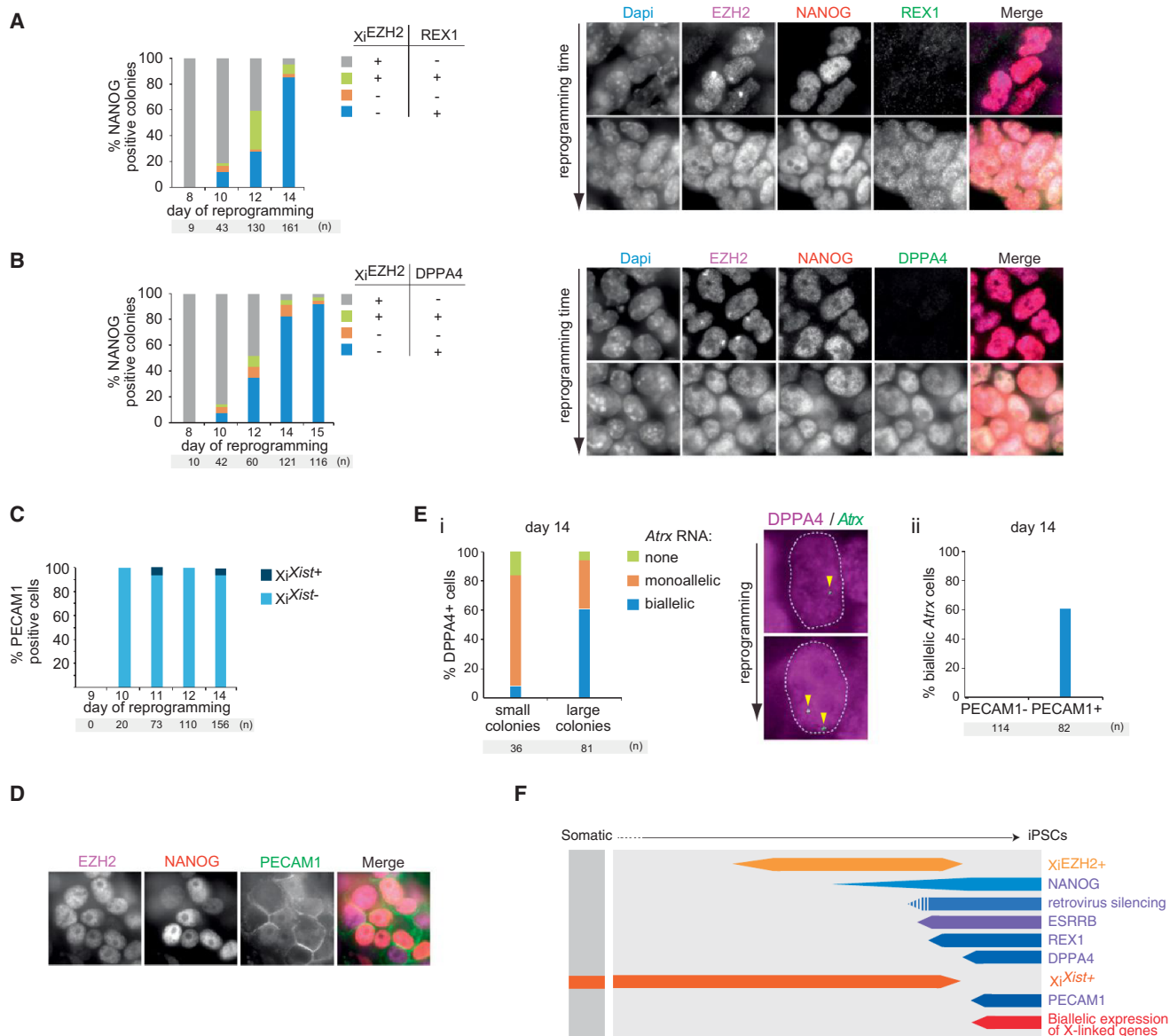


Figure 3. Xi Features in Relation to the Sequential Expression of Pluripotency Factors during Reprogramming

(A) Quantitation of an immunostaining analysis of NANOG, Xi^{EZH2+} , and REX1, presenting the proportion of NANOG+ colonies with Xi^{EZH2+} and/or REX1 expression at indicated time points. Right, representative immunostaining images for EZH2 (magenta in the merge), NANOG (red), and REX1 (green).

(B) As (A) for NANOG, Xi^{EZH2+} , and DPPA4.

(C) Quantitation of an immunoFISH analysis for PECAM1 and *Xist* RNA displaying the proportion of PECAM1+ cells with Xi^{Xist+} .

(D) Representative immunostaining image for EZH2 (magenta in the merge), NANOG (red), and PECAM1 (green), demonstrating the absence of Xi^{EZH2+} in PECAM1+ cells.

(E) (i) Quantitation of mono- and biallelic expression of the X-linked gene *Atrx* within cells of small (<12 cells) and large (>20 cells) DPPA4+ colonies and representative images from the DPPA4 (magenta) and *Atrx* (green) immunoFISH staining. Arrowheads indicate *Atrx* nascent transcription signals. (ii) Quantification of biallelic *Atrx* expression in PECAM1+/- cells.

(F) Summary of reprogramming stages identified in this figure as in Figure 2E.

See also Figure S3.

expression did not alter the efficiency by which ESRRB+ colonies appeared but resulted in a decrease in XCR within NANOG+ cells, as measured by the extent of biallelic *Atrx* expression (Figures 5D–5F). Thus, *Xist* silencing at the end of reprogramming is necessary for XCR.

To determine whether XCR depends solely on *Xist* repression, we asked whether *Xist* deletion leads to precocious activation of the Xi. Specifically, we deleted *Xist* early in the reprogramming process using female MEFs homozygous for a conditional (2lox) *Xist* allele (Csankovszki et al., 2001), which also carried a

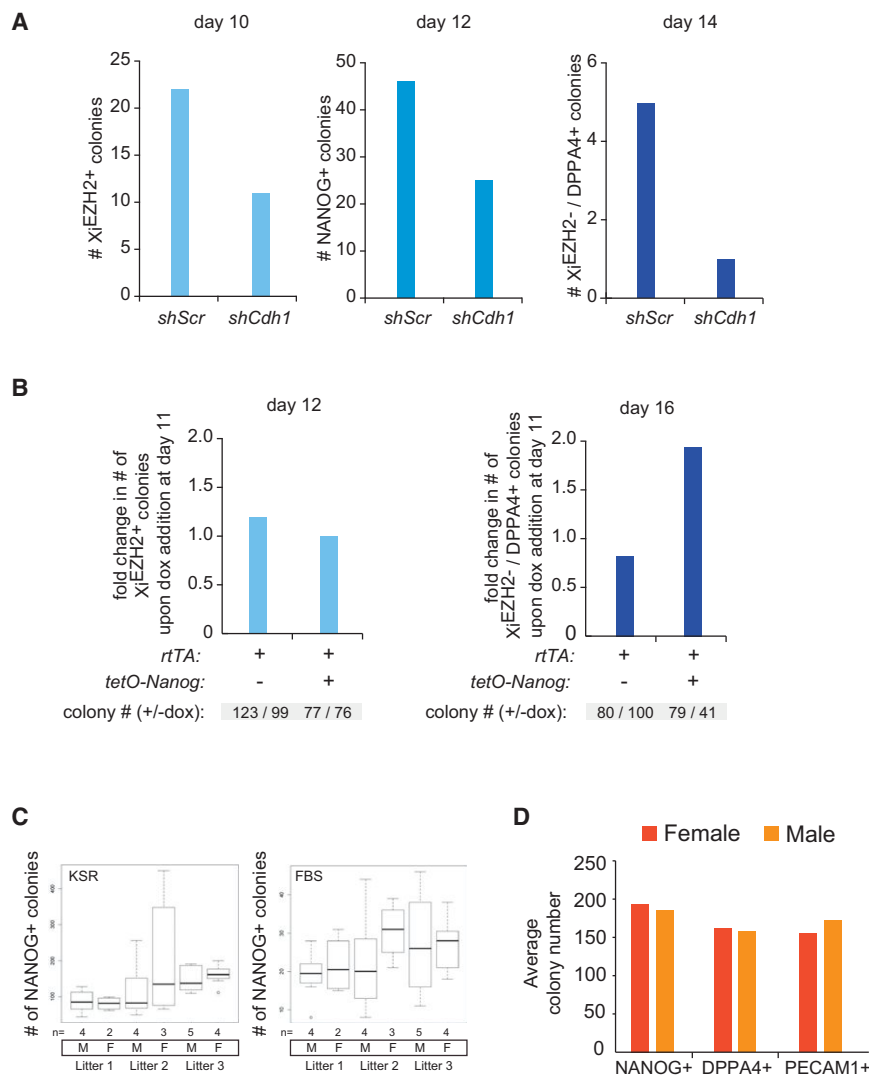


Figure 4. *Cdh1* and *Nanog* Modulate the Efficiency of Reprogramming and Dynamics of Xi Hallmarks, whereas XCR Does Not Represent a Reprogramming Barrier

(A) Number of XiEZH2+, NANOG+, and DPPA4+/XiEZH2- colonies obtained when *Cdh1* is knocked down by shRNAs (*shCdh1*) throughout reprogramming, compared to scrambled shRNA (*shScr*) reprogramming experiments.

(B) Reprogramming experiments with female MEFs carrying *rtTA* only or *rtTA* and the *tetO-Nanog* allele, and with and without dox addition at day 11. The number of XiEZH2+ and DPPA4+/XiEZH2- colonies was determined at day 12 and 16 of reprogramming, respectively, and was plotted as fold change between dox relative to no dox treatment per cell line. Note that XiEZH2+ counts are similar at day 12, whereas DPPA4+/XiEZH2- counts differ at day 14 when *Nanog* is overexpressed.

(C) Comparison of reprogramming efficiency between male (M) and female (F) MEFs. Box plots depict the number of NANOG+ colonies for reprogramming experiments in KSR and FBS media, respectively, with male and female MEFs, at day 14 and day 25 of reprogramming, respectively. MEFs are grouped by litter and the number of male and female MEF populations per litter is given (n). Whiskers demarcate the minimum and maximum of the data.

(D) Quantitation of different late reprogramming stages for MEFs isolated from seven female and seven male embryos, as judged by the number of NANOG+, DPPA4+, or PECAM1+ colonies at day 14.

See also Figure S4.

dox-inducible Cre recombinase (Figures 5G and 5H). *Xist* ablation had no effect on the efficiency with which NANOG+ colonies were generated (Figure 5I) and, surprisingly, did not alter XCR kinetics (Figure 5J). Therefore, *Xist* repression is necessary, but not sufficient, for XCR to occur, indicating the existence of other mechanisms that can maintain Xi silencing throughout reprogramming and even initially in NANOG+ cells.

High Persistence of Xi DNA Methylation during Reprogramming

We considered the possibility that DNA methylation could maintain the silent state of the Xi during reprogramming in the absence of *Xist*. DNA methylation at CpG islands of the Xi arises late in the sequence of epigenetic changes on the Xi during development (Gendrel et al., 2012). We examined the DNA methylation pattern of X-linked genes in SSEA1- and SSEA1+ subpopulations, isolated from day 9 reprogramming cultures, representing cell populations with different reprogramming capabilities (Figures 1L and 1M; Polo et al., 2012). Traditional bisul-

Nanog promoter region, methylated at an intermediate level in MEFs and SSEA1- cells, displayed demethylation characteristic of pluripotent cells already in SSEA1+ cells. These findings suggested a differential persistence of the methylation mark between *Nanog* and Xi-linked genes.

To determine the DNA methylation status along the entire X chromosome, we employed reduced representative bisulfite sequencing (Meissner et al., 2008), which provides genome-scale single-base-resolution maps of DNA methylation. For this analysis, we additionally included early passage female iPSCs, as well as male MEFs and male ESCs for comparison. CpG islands on the Xa in male cells were hypomethylated to the same degree as those on autosomes in male or female cells (Figure 6B). By contrast, in female MEFs, CpG islands across the X chromosome showed an average of 20%–50% methylation, which is consistent with an Xi-specific methylation signature. This pattern was present in both SSEA1- and SSEA1+ subpopulations but absent in early-passage female iPSCs and female ESCs (Figure 6B). A similar result was obtained for CpG-island

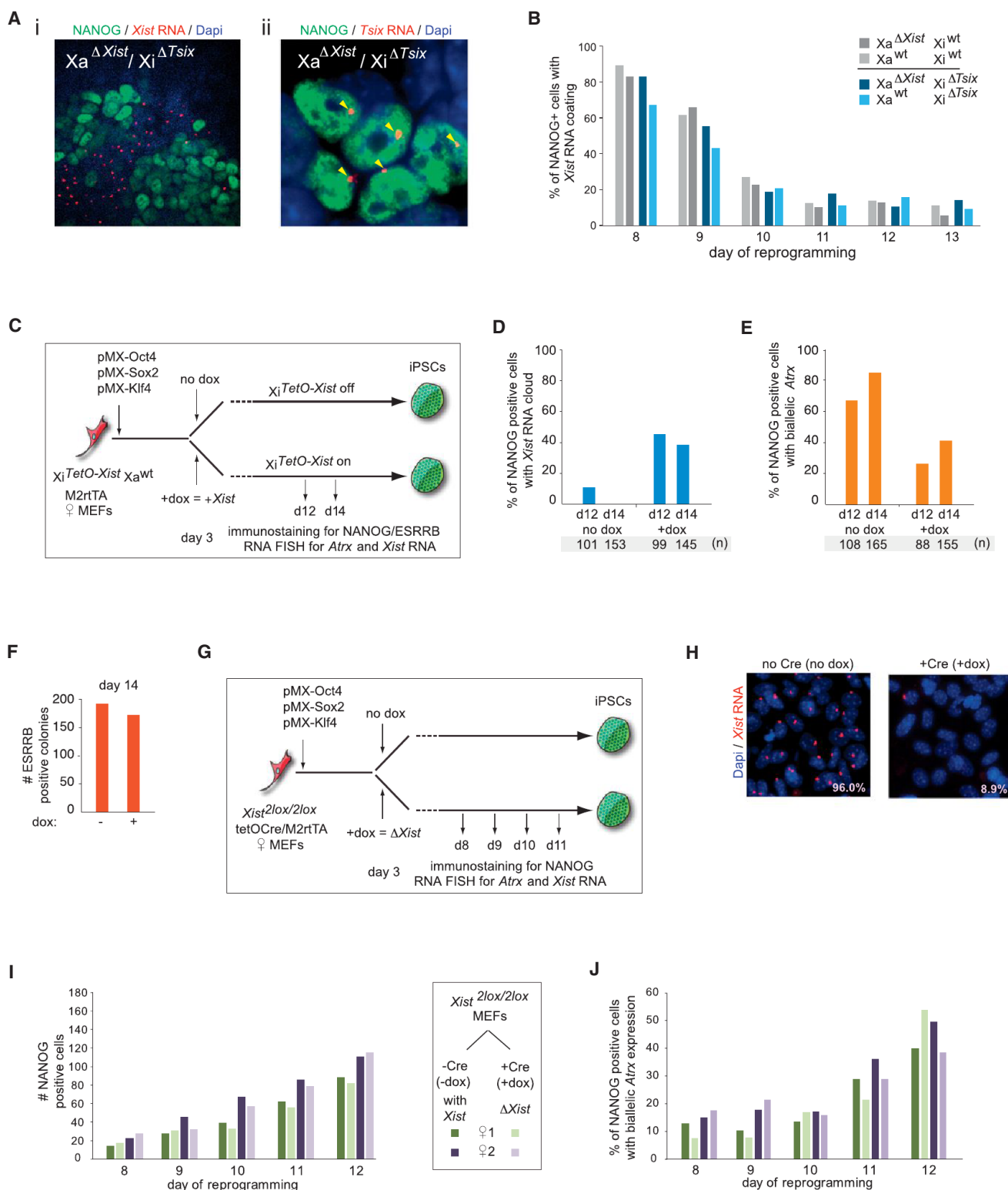


Figure 5. *Xist* Silencing Is Necessary, but Not Sufficient, for XCR

(A) (i) Representative image of an immunoFISH analysis for NANOG (green) and *Xist* RNA (red) at day 14 of reprogramming with MEFs carrying a deletion of *Tsix* on the Xi and of *Xist* on the Xa, illustrating that NANOG+ cells lose *Xist* RNA accumulation on the Xi even in the absence of *Tsix* on the Xi. (ii) iPSCs derived from the experiment in (i) were stained for NANOG (green) and *Tsix* RNA (red), confirming monoallelic expression of *Tsix* due to deletion on one X chromosome (arrowheads).

(legend continued on next page)

shores, high and low CpG-containing promoters (Figure S5A). These results indicate that DNA methylation established on the Xi late during differentiation (Gendrel et al., 2012) is preserved on the Xi until very late in reprogramming. The persistence of Xi-DNA methylation in reprogramming is *Xist* independent (Figures S5B and S5C), supporting the hypothesis that this Xi mark could maintain the silent state of the Xi until late in reprogramming, even when *Xist* is experimentally deleted. Because Xi-DNA methylation is not yet reversed when *Nanog* (Figure 6A) and many other ESC-specific enhancer elements have already become demethylated (V.P., R.K., C. Chronis, A.M., and K.P., unpublished data), we conclude that this Xi mark has a remarkable stability during reprogramming.

***Xist* RNA and DNA Methylation Both Maintain Xi Silencing Throughout Reprogramming**

We determined whether XCR is mechanistically linked to the loss of both *Xist* and DNA methylation by deleting *Xist* and inhibiting *Dnmt1*, the maintenance DNA methyltransferase, during the late phase of reprogramming. The block of *Dnmt1* activity and loss of DNA methylation within Xi-linked CpG islands was confirmed (Figures S5D and S5E). In reprogramming cultures in which *Xist* on the Xi was experimentally deleted, 23% of NANOG⁺ cells displayed biallelic expression of the X-linked gene *Atrx* upon brief *Dnmt1* depletion, and this proportion was more than doubled when *Dnmt1* knockdown was combined with 5AzadC treatment to enhance DNA demethylation (Figures 6C and S5E). XCR was not detected at this time point in NANOG⁺ cells in control reprogramming cultures (Figure 6C). Importantly, we found that inhibition of DNA methylation only enhances XCR in NANOG⁺ cells in the absence of *Xist*, but not in its presence (Figure 6D). This finding also excludes the possibility that the acceleration of XCR upon inhibition of DNA methylation is simply due to faster overall reprogramming. We conclude that *Xist* RNA is able to maintain the Xi when DNA methylation is reduced and that DNA methylation is sufficient for Xi maintenance in the absence of *Xist*. Therefore, both DNA demethylation and *Xist* silencing are required for XCR late in reprogramming and occur downstream to the reactivation and demethylation of *Nanog* (Figure 6E).

***Tet1*, *Tet2*, and High Global 5hmC Levels Are Dispensable for XCR**

Given the implication of conversion of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) in DNA demethylation processes (Wu and Zhang, 2014), we defined the Xi-specific and

global dynamics of 5hmC during reprogramming. We found a striking increase in the global 5hmC level, specifically in those cells that globally upregulate EZH2 and gain Xi^{EZH2+} (Figures 7A–7C). Global upregulation of 5hmC also took place in male reprogramming cultures and in the absence of Vitamin C (Figure S6A), indicating that this epigenetic remodeling event is intrinsic to reprogramming across different culture conditions and sex chromosome content. Despite overall elevated 5hmC levels, this mark was depleted on the Xi in Xi^{EZH2+} cells (Figures 7D and 7E). Thus, during reprogramming, cells start off with low levels of 5hmC and EZH2 and then increase 5hmC and EZH2 downstream of MET, with PRC2 accumulating on the Xi and 5hmC remaining excluded from the Xi, all of which precedes the reactivation of pluripotency genes and transition to a pluripotent state with XCR devoid of Xi^{PRC2+} and 5hmC Xi exclusion.

Given the dynamics of 5hmC, we tested the requirement of *Tet1* and *Tet2* for XCR using female MEFs carrying *Tet1* knockout (*Tet1*^{−/−}) and *Tet2* conditional (*Tet2*^{2lox/2lox}) alleles, in which genetic deletion of *Tet2* could be induced by addition of Cre-expressing adenoviruses (AdCre) (Figure S6B). Strikingly, genetic ablation of both *Tet1* and *Tet2*, but not that of either *Tet1* or *Tet2* individually, prevented the global induction of 5hmC during reprogramming (Figures 7F, S6C, and S6D). Importantly, *Tet1/Tet2* double knockout and absence of global 5hmC did not affect the upregulation of nuclear EZH2 and occurrence of Xi^{EZH2+}, nor the efficiency with which NANOG⁺ colonies were obtained, nor the activation of the late pluripotency marker PE-CAM1 and XCR (Figures 7G, 7H, and S6E–S6G). Reprogramming experiments with ablation of either *Tet1* or *Tet2* resulted in similar results, and the resulting iPSCs contributed to chimeras and were effectively demethylated at *cis*-regulatory regions of the *Pou5f1* (*Oct4*) gene and Xi-linked promoters (Figures S6H–S6K and S7A–S7D). Additional shRNA-mediated depletion of *Tet3* transcripts in pre-iPSCs also carrying the *Tet1* and *Tet2* genetic deletion still enabled XCR (Figures S7E–S7K). We conclude that *Tet1* and *Tet2* and the global increase in 5hmC nuclear levels are dispensable for XCR and the transition through the reprogramming hierarchy that we have established.

DISCUSSION

A dramatic reorganization of the epigenome occurs during the reprogramming of somatic cells to iPSCs. Our findings demonstrate that changes in global and Xi-specific chromatin states, noncoding RNA expression, and pluripotency-associated factor

(B) Kinetics of *Xist* RNA loss in the absence of *Tsix* on the Xi. Proportion of NANOG⁺ cells with Xi^{Xist+} in reprogramming time courses performed with MEFs with (gray bars) and without (blue bars) *Tsix* on the Xi. Just like *Tsix* deletion, the additional deletion of *Xist* on the Xa (dark versus lighter bars) does not affect the kinetics of *Xist* RNA loss in NANOG⁺ cells.

(C) Diagram of *Xist* overexpression reprogramming experiments using MEFs in which the promoter of *Xist* on the Xi is replaced with a tet-inducible promoter.

(D) Proportion of NANOG⁺ cells with Xi^{Xist+} in reprogramming cultures described in (C) with and without ectopic *Xist* induction conditions (+/−dox), based on immunoFISH analysis.

(E) As (D), but for NANOG⁺ cells with biallelic *Atrx* expression.

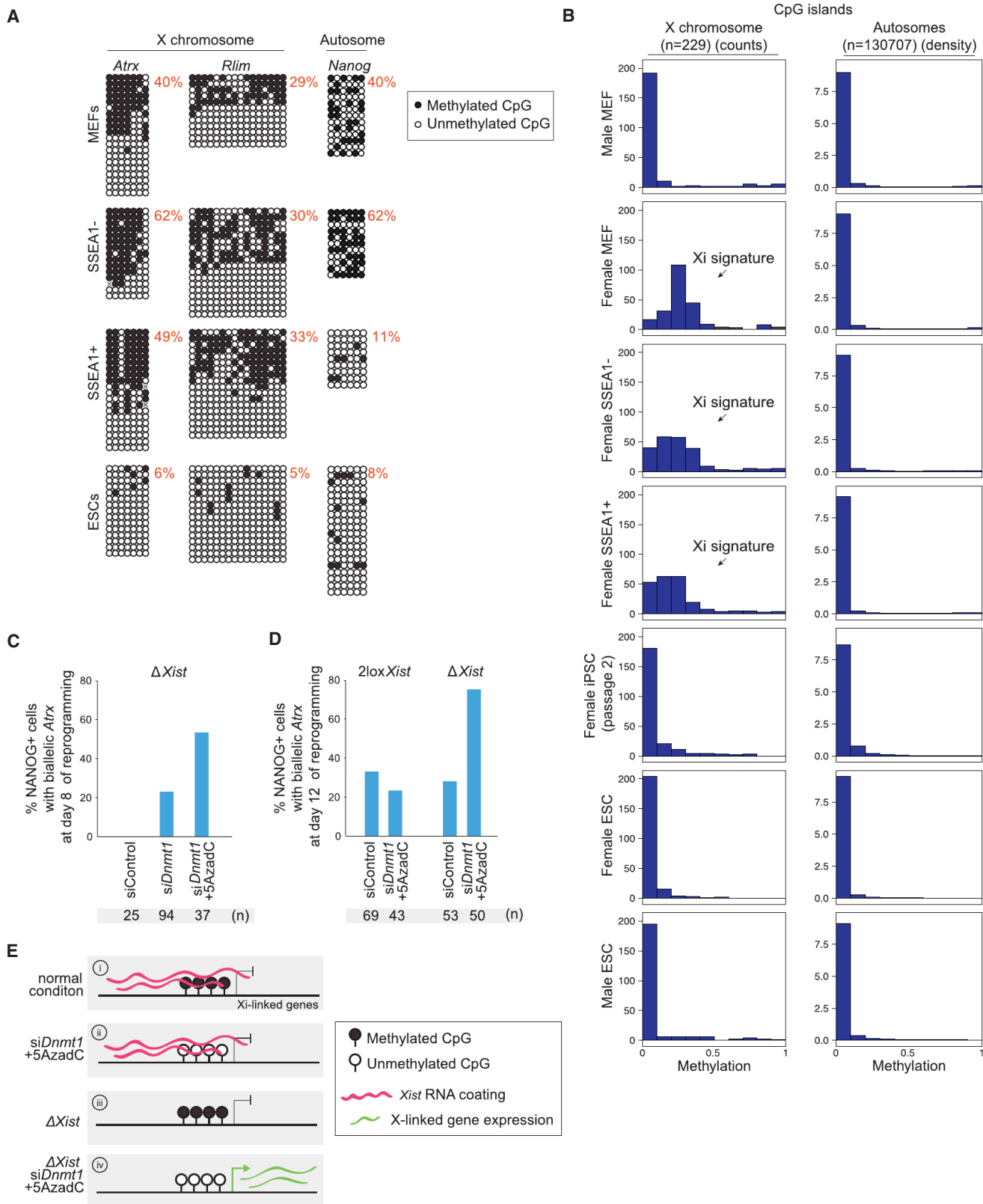
(F) As (D), but for the number of ESRRB⁺ colonies.

(G) Diagram of the *Xist* deletion reprogramming experiments with female conditional *Xist* MEFs.

(H) *Xist* RNA FISH for MEFs described in (G) under control (−dox) and +dox conditions, the latter leading to *Xist* RNA loss in the majority of cells.

(I) Number of NANOG⁺ colonies at various time points of reprogramming for the experiment described in (G) under control (no dox/−Cre) and the *Xist* deletion (+dox/+Cre) conditions.

(J) As in (I), but quantitation of NANOG⁺ cells with biallelic *Atrx* expression based on immunoFISH analysis.



(legend on next page)

expression are highly reproducible and reveal the existence of a multitude of epigenetic steps that occur in a defined sequence throughout the reprogramming process (Figure 7I, i). For instance, focusing only on $\text{Xi}^{\text{EZH2}+}$ dynamics relative to CDH1 and NANOG expression, transition through four steps can be defined: (1) $\text{CDH1}+/\text{Xi}^{\text{EZH2}-}/\text{NANOG}-$; (2) $\text{CDH1}+/\text{Xi}^{\text{EZH2}+}/\text{NANOG}-$; (3) $\text{CDH1}+/\text{Xi}^{\text{EZH2}+}/\text{NANOG}+$; and (4) $\text{CDH1}+/\text{Xi}^{\text{EZH2}-}/\text{NANOG}+$ (Figure 7I, ii). These stages are likely going to be generally applicable to female cells and not cell-type specific as the Xi enrichment of PRC2 is also expected to occur in epithelial cells at an intermediate step of reprogramming. The relocalization of EZH2 (PRC2) and its cofactor JARID2 to the Xi, along with global increases in PRC2, macroH2A1, and 5hmC downstream of MET and upstream of NANOG expression, indicate that major changes in chromatin structure take place in cells undergoing reprogramming, before pluripotency is reached.

Compared to the establishment of Xi features during differentiation, we find that these have different propensities for reversal during reprogramming. Whereas $\text{Xi}^{\text{EZH2}+}$ and the activation of *Tsix* from the Xa and Xi take place in an apparent reverse order of the developmental XCI program, macroH2A1 and DNA methylation, both associated with the differentiated state and resistance to reprogramming (Pasque et al., 2012; Mikkelsen et al., 2008), are reversed on the Xi only very late in reprogramming, despite the fact that they are established on the Xi late in differentiation. Similarly, the activation of Xi-linked genes during reprogramming only occurs after *Xist* RNA loss, even though *Xist* RNA coating precedes silencing of the X chromosome during differentiation. Thus, based on a subset of Xi hallmarks, reprogramming proceeds in a manner that would be expected for developmental reversal, indicating progressive dedifferentiation. However, based on another set of marks, cells undergoing reprogramming remain epigenetically distinct from those traversing differentiation. Thus, during reprogramming, certain epigenetic features follow the differentiation state of the cell, whereas others are uncoupled from this regulation.

During differentiation, *Xist* is required to initiate XCI, and its experimental silencing in the first days after the establishment of the Xi leads to immediate reactivation of the X chromosome (Wutz and Jaenisch, 2000). However, later in differentiation, *Xist* can be deleted from the Xi without dramatically affecting the stability of the silent chromosome, which, at this point, is thought to be maintained through the action of multiple repressive chromatin pathways (Csankovszki et al., 2001). In contrast to a recent study that used a different system to reduce *Xist*

expression (Chen et al., 2014), we made the surprising observation that *Xist* ablation on the Xi does not alter the kinetics of XCR during reprogramming. Our result indicates that the extended, several-day-long window of *Xist* dependency of silencing seen during the initiation of XCI in differentiation (Wutz and Jaenisch, 2000) is not re-established during reprogramming. Silencing of the Xi in the absence of *Xist* remains stable until the very end of reprogramming because it is functionally maintained by DNA methylation, which has an extraordinarily high persistence on the Xi during reprogramming and is only erased after the pluripotency factor *Nanog* is already demethylated. Notably, the experimental interference with DNA methylation alone does not lead to precocious XCR, indicating that *Xist* RNA also actively contributes to the silencing of the Xi late in reprogramming. In agreement with this, we also discovered that forced *Xist* expression prevents XCR during reprogramming. Therefore, XCR requires loss of both *Xist* RNA and DNA methylation at the end of the reprogramming process. Because both events take place only late during hierarchical pluripotency-associated gene activation, these ensure that XCR only occurs in cells that establish faithful pluripotency (Figure 7I, iii). Accordingly, a block early in the pluripotency hierarchy blocks XCR (Figure S4). Notably, the pluripotency factor PRDM14 has been reported to be required for XCR during reprogramming (Payer et al., 2013), but whether *Prdm14* deletion blocks the reprogramming process at a stage prior to XCR needs to be resolved to understand its specific role in XCR.

The generation of 5hmC by Tet proteins has been suggested to play important roles during reprogramming to iPSCs and potentially mediates DNA demethylation through active and passive mechanisms (Hu et al., 2014; Wu and Zhang, 2014). Our findings reveal that *Tet1*, *Tet2*, and global 5hmC are dispensable for XCR. This raises the question of which DNA demethylation pathway, either active or passive, leads to XCR. We posit that loss of DNA methylation on the Xi during reprogramming likely occurs in a synchronous manner across the entire chromosome, requiring a mechanism that can act across a large number of CpG islands in a relatively short time frame. We expect that the characterization of the Xi DNA demethylation event will yield critical insights into mechanisms that control the final stages of reprogramming.

Notably, during reprogramming by SCNT, developmental defects are caused by misregulation of the XCI system, particularly due to ectopic *Xist* expression from the Xa (Inoue et al., 2010). By contrast, our data indicate that, during reprogramming to iPSCs,

Figure 6. Analysis of DNA Methylation on the X Chromosome during Reprogramming

(A) Bisulfite PCR analysis of the promoter regions of the X-linked genes *Atrx* and *Rlim* and of *Nanog* in female MEFs, ESCs, and day 9 SSEA1^{-/+} reprogramming intermediates. Black circles indicate methylated CpGs, and open circles indicate unmethylated CpGs. The proportion of methylated CpGs is given. For MEFs and SSEA1^{+/+} cells, hemimethylation represents Xi methylation.

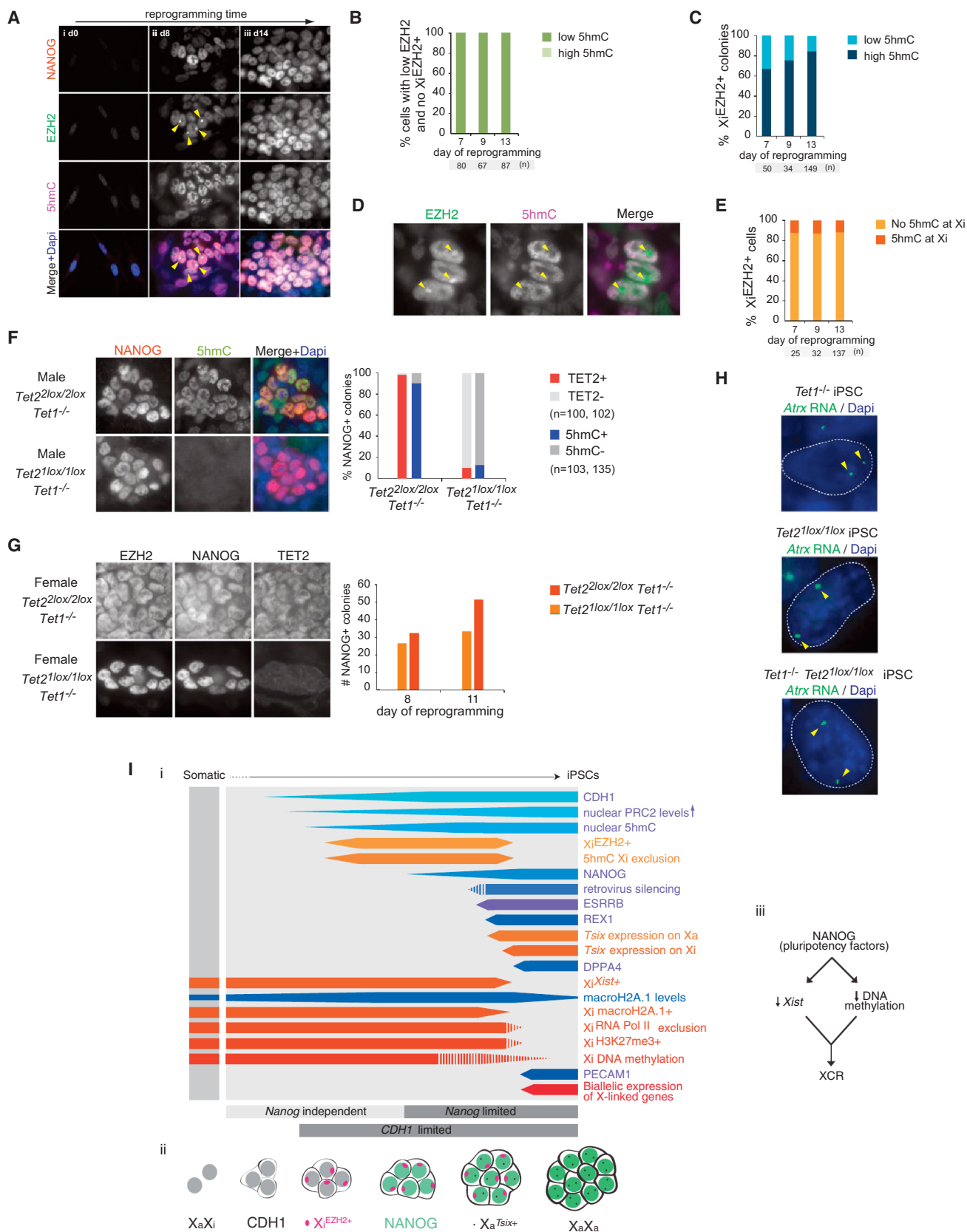
(B) Histograms showing the distribution of methylation levels across CpG islands on the X chromosome and autosomes in indicated cell types based on RRBS data (n, number of CpG islands). The arrow indicates the Xi-specific DNA methylation signature.

(C) Proportion of NANOG⁺ cells with biallelic *Atrx* expression based on immunoFISH analysis at day 8 of reprogramming with *2lox/2lox Xist* MEFs in which *Xist* was deleted by activation of the dox-inducible Cre-recombinase, and siControl, si*Dnmt1*, and si*Dnmt1* plus 5AzadC, respectively, the latter were added at day 5. All NANOG⁺ cells present in the culture were counted (n).

(D) Similar to (C), except that *Xist* deletion was performed only in half of the reprogramming culture, and siControl or si*Dnmt1*+5AzadC were applied on day 5 and day 8. At day 12, all NANOG⁺ cells present in the culture were assessed for biallelic *Atrx* expression.

(E) Summary of the role of *Xist* RNA and DNA methylation in the control of gene silencing on the Xi during reprogramming.

See also Figure S5.



(legend on next page)

the reactivation of the Xi (Figure 4) or ectopic XCI on the Xa (Figures 5B and S7L) does not seem to act as barriers to reprogramming, pointing to mechanistic differences between transcription factor- and oocyte-induced somatic cell reprogramming. Furthermore, in contrast to our findings in iPSC reprogramming, the activation of X-linked genes during XCR in preimplantation development occurs in the presence of *Xist* (Williams et al., 2011).

Importantly, our study defines many sequential reprogramming steps, extending previous reports based on gene expression studies that identified a limited number of reprogramming stages (Parchem et al., 2014; O'Malley et al., 2013; Buganim et al., 2012; Polo et al., 2012). We propose that the global epigenetic state of cells as they reprogram to iPSCs, and that of the Xi, is less variable than transcriptional states. However, our data do not exclude stochastic gene expression differences in cells with the same epigenetic state. One advantage of our analyses is that the stage of any cell in a reprogramming culture can be easily assessed, taking into account criteria such as colony growth and positional information of cells, as well as protein levels and subcellular localization. Notably, although most of our analyses focused on the female-specific XCR process, our work led to the identification of many reprogramming stages that are also applicable to male reprogramming (Figure 7I, i). For example, the global increase in EZH2 and 5hmC levels that occurs in both female and male cells was uncovered during our analysis of the localization of these marks on the Xi in female cells.

Our study provides an easily applicable platform for assaying the effects of interference with intrinsic and extrinsic factors on the stages of reprogramming and on the transitions between them. Additionally, we anticipate that the analysis of the transcriptome and other epigenetic features such as DNA methylation in the multiple reprogramming intermediates that we have identified will reveal insights into reprogramming. Another task ahead remains the continuous imaging of the transitions between the reprogramming steps identified here to quantitatively model the reprogramming process.

In conclusion, our comprehensive study yields insights into XCR and provides unprecedented details on the epigenetic dy-

namics of somatic cell reprogramming to induced pluripotency, establishing a valuable foundation exploitable for many applications, including staging of reprogramming cultures, isolation of intermediates, and to uncover mechanistically how cells transition toward pluripotency.

EXPERIMENTAL PROCEDURES

Reprogramming Experiments and Time Courses

Reprogramming was carried out using cells derived from reprogrammable mice or directly infected with retroviruses encoding *Oct4*, *Sox2*, and *Klf4*, as described in detail in the Extended Experimental Procedures. For time course analyses, reprogramming cultures on 22 × 22 mm gelatinized glass coverslips were fixed every other day, usually from day 6 to day 14, before carrying out immunostaining and RNA FISH analyses.

Flow Cytometry

Flow cytometry for SSEA1 and CDH1 was done starting from large reprogramming cultures using methods previously reported (Stadtfield et al., 2008) with modifications described in the Extended Experimental Procedures.

Immunostaining and RNA FISH

Immunostainings and RNA FISH were carried out on 22 × 22 coverslips obtained from reprogramming cultures and as described previously (Maherali et al., 2007). Details are given in the Extended Experimental Procedures.

Bisulfite Analysis

Bisulfite-converted DNA was subjected to RRBS or analyzed by PCR as detailed in Table S1. Details are given in the Extended Experimental Procedures.

Data Analyses

See the Extended Experimental Procedures.

ACCESSION NUMBERS

The GEO accession number for the RRBS data reported in this paper is GSE58109.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.040>.

Figure 7. *Tet1* and *Tet2* and Global 5hmC Are Dispensable for XCR

- (A) Representative immunostaining images for different patterns of NANOG (red in merge), EZH2 (green), and 5hmC (magenta) arising at indicated days of reprogramming. Arrowheads indicate Xi^{EZH2+} .
- (B) Proportion of cells with low nuclear EZH2 levels and no Xi^{EZH2+} that display either low or high nuclear levels of 5hmC at indicated time points.
- (C) Proportion of Xi^{EZH2+} colonies that display either low or high 5hmC at indicated time points.
- (D) Representative immunostaining image for EZH2 (green in merge) and 5hmC (magenta) in the Xi^{EZH2+} reprogramming intermediate. 5hmC Xi exclusion is indicated by arrowheads.
- (E) Proportion of Xi^{EZH2+} cells that display 5hmC Xi exclusion (Xi^{5hmC-}) at indicated time points.
- (F) Representative immunostaining images of male $Tet2^{2lox/2lox}Tet1^{-/-}$ reprogramming cultures infected with Ad5 (top) or AdCre (bottom) adenoviruses, stained for NANOG (red in merge) and 5hmC (green) at day 14 of reprogramming. AdCre induces *Tet2* deletion ($Tet2^{1lox/1lox}Tet1^{-/-}$). The graph gives the proportion of NANOG+ colonies positive for 5hmC and TET2, respectively, at day 14 based on immunostaining. The absence of the TET2 signal in NANOG+ cells confirms effective *Tet2* deletion. Loss of both *Tet1* and *Tet2* leads to loss of the 5hmC immunostaining signal (loss of global 5hmC).
- (G) As in (F), except for female $Tet2^{2lox/2lox}Tet1^{-/-}$ and $Tet2^{1lox/1lox}Tet1^{-/-}$ reprogramming cultures, immunostained for EZH2, NANOG, and TET2. The number of NANOG+ colonies at indicated time points in these cultures is given in the graph.
- (H) RNA FISH for *Atrx* nascent transcription on female $Tet1^{-/-}$, $Tet2^{1lox/1lox}$ and $Tet2^{1lox/1lox}Tet1^{-/-}$ iPSCs. Arrowheads indicate the biallelic *Atrx* signal.
- (I) Stages of XCR and somatic cell reprogramming to induced pluripotency. Our view of the stages leading to XCR and the induction of pluripotency, shown as described in Figures 1I and 2E. Female-specific events are shown in orange/red, and those occurring in both female and male cells are shown in blue. With the exception of retroviral silencing in male reprogramming, all results presented are based on experimental evidence in both female and male reprogramming. See also Figures S6 and S7.

AUTHOR CONTRIBUTIONS

V.P., J.T., and K.P. designed experiments; V.P., J.T., R.K., M.U., A.S.D., D.C., B.P., S.P., R.M., and K.P. performed experiments; V.P., J.T., R.K., M.U., A.S.D., D.C., B.P., G.B., S.P., R.M., A.M., and K.P. analyzed data; R.H. helped with experiments; R.S., T.S., and T.T. generated reagents; and V.P. and K.P. wrote the manuscript with edits from J.T., R.K., B.P., and A.M.

ACKNOWLEDGMENTS

We are grateful to Drs. Nakamura, Xu, Eng, Berk, Silva, Pei, and Besser for providing reagents; Dr. Chronis for help with bioinformatics; Dr. Ch'ng and A. Sahakyan for help with microscopy and western blot analysis; F. Codrea and J. Scholes at the Broad Stem Cell Center FACS Core for help; and Dr. Lowry and members of the Plath lab for advice and critical reading of the manuscript. V.P., S.P., and A.S.D. are supported by CIRM Training Grants TG2-01169 and TB1-01183; J.T. is supported by a fellowship of the UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research; A.M. is a NYSCF Robertson Investigator and is supported by P01GM099117; K.P. is supported by the UCLA Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research and NIH P01 GM099134 and CIRM (RN1-00564); D.C. is supported by the NIH Ruth L. Kirschstein National Research Service Award (GM007185); G.B. is supported by the Whitcome Pre-doctoral Training Program and a UCLA Dissertation Year Fellowship; M.U. is supported by R25GM055052; R.H. is supported by an NIH Training Grant (5T32AI060567-07) and the UCLA Graduate Division Dissertation Year Fellowship; and R.S. is supported by the UCLA Graduate Division Dissertation Year Fellowship.

Received: June 9, 2014

Revised: September 30, 2014

Accepted: November 12, 2014

Published: December 18, 2014

REFERENCES

- Apostolou, E., and Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. *Nature* 502, 462–471.
- Barakat, T.S., and Gribnau, J. (2010). X chromosome inactivation and embryonic stem cells. *Adv. Exp. Med. Biol.* 695, 132–154.
- Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 150, 1209–1222.
- Buganim, Y., Faddah, D.A., and Jaenisch, R. (2013). Mechanisms and models of somatic cell reprogramming. *Nat. Rev. Genet.* 14, 427–439.
- Carter, A.C., Davis-Dusenbery, B.N., Koszka, K., Ichida, J.K., and Eggan, K. (2014). Nanog-independent reprogramming to iPSCs with canonical factors. *Stem Cell Reports* 2, 119–126.
- Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* 20, 2223–2237.
- Chen, Q., Gao, S., He, W., Kou, X., Zhao, Y., Wang, H., and Gao, S. (2014). Xist repression shows time-dependent effects on the reprogramming of female somatic cells to induced pluripotent stem cells. *Stem Cells* 32, 2642–2656.
- Chow, J., and Heard, E. (2009). X inactivation and the complexities of silencing a sex chromosome. *Curr. Opin. Cell Biol.* 21, 359–366.
- Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J. Cell Biol.* 153, 773–784.
- da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C., et al. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol. Cell* 53, 301–316.
- Eggan, K., Akutsu, H., Hochedlinger, K., Rideout, W., 3rd, Yanagimachi, R., and Jaenisch, R. (2000). X-Chromosome inactivation in cloned mouse embryos. *Science* 290, 1578–1581.
- Gendrel, A.-V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., et al. (2012). Smcnd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev. Cell* 23, 265–279.
- Golipour, A., David, L., Liu, Y., Jayakumar, G., Hirsch, C.L., Trcka, D., and Wrana, J.L. (2012). A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Cell Stem Cell* 11, 769–782.
- Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creighton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* 462, 595–601.
- Hu, X., Zhang, L., Mao, S.-Q., Li, Z., Chen, J., Zhang, R.-R., Wu, H.-P., Gao, J., Guo, F., Liu, W., et al. (2014). Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* 14, 512–522.
- Inoue, K., Kohda, T., Sugimoto, M., Sado, T., Ogonuki, N., Matoba, S., Shiura, H., Ikeda, R., Mochida, K., Fujii, T., et al. (2010). Impeding Xist expression from the active X chromosome improves mouse somatic cell nuclear transfer. *Science* 330, 496–499.
- Lee, J.T., and Bartolomei, M.S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* 152, 1308–1323.
- Lee, J.T., and Lu, N. (1999). Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* 99, 47–57.
- Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* 7, 51–63.
- Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* 1, 55–70.
- Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766–770.
- Mermoud, J.E., Costanzi, C., Pehrson, J.R., and Brockdorff, N. (1999). Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. *J. Cell Biol.* 147, 1399–1408.
- Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. *Nature* 454, 49–55.
- O'Malley, J., Skylaki, S., Iwabuchi, K.A., Chantzoura, E., Ruetz, T., Johnsson, A., Tomlinson, S.R., Linnarsson, S., and Kaji, K. (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature* 499, 88–91.
- Parchem, R.J., Ye, J., Judson, R.L., LaRussa, M.F., Krishnakumar, R., Blelloch, A., Oldham, M.C., and Blelloch, R. (2014). Two miRNA clusters reveal alternative paths in late-stage reprogramming. *Cell Stem Cell* 14, 617–631.
- Pasque, V., Radziszewska, A., Gillich, A., Halley-Stott, R.P., Panamara, M., Zernicka-Goetz, M., Surani, M.A., and Silva, J.C.R. (2012). Histone variant macroH2A marks embryonic differentiation in vivo and acts as an epigenetic barrier to induced pluripotency. *J. Cell Sci.* 125, 6094–6104.
- Payer, B., Rosenberg, M., Yamaji, M., Yabuta, Y., Koyanagi-Aoi, M., Hayashi, K., Yamanaka, S., Saitou, M., and Lee, J.T. (2013). Tsix RNA and the germline factor, PRDM14, link X reactivation and stem cell reprogramming. *Mol. Cell* 52, 805–818.
- Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300, 131–135.
- Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A

molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* 151, 1617–1632.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.-K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* 7, 64–77.

Schwarz, B.A., Bar-Nur, O., Silva, J.C.R., and Hochedlinger, K. (2014). Nanog is dispensable for the generation of induced pluripotent stem cells. *Curr. Biol.* 24, 347–350.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H.F.M., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev. Cell* 4, 481–495.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. *Cell* 138, 722–737.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* 2, 230–240.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.

Williams, L.H., Kalantry, S., Starmer, J., and Magnuson, T. (2011). Transcription precedes loss of Xist coating and depletion of H3K27me3 during X-chromosome reprogramming in the mouse inner cell mass. *Development* 138, 2049–2057.

Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* 156, 45–68.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* 5, 695–705.

High-Resolution Sequencing and Modeling Identifies Distinct Dynamic RNA Regulatory Strategies

Michal Rabani,^{1,2,12} Raktima Raychowdhury,¹ Marko Jovanovic,¹ Michael Rooney,^{1,3} Deborah J. Stumpo,⁴ Andrea Pauli,⁵ Nir Hacohen,¹ Alexander F. Schier,^{1,5,6} Perry J. Blackshear,^{4,7,11} Nir Friedman,^{8,11} Ido Amit,^{9,11} and Aviv Regev^{1,10,11,*}

¹The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

²Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02140, USA

³Harvard/MIT Division of Health Sciences and Technology, Cambridge, MA 02141, USA

⁴The Laboratory of Signal Transduction, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA

⁵Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

⁶FAS Center for Systems Biology, Harvard University, Cambridge, MA 02138, USA

⁷Departments of Medicine and Biochemistry, Duke University Medical Center, Durham, NC 27710, USA

⁸School of Computer Science and Institute of Life Sciences, Hebrew University, Jerusalem 91904, Israel

⁹Department of Immunology, Weizmann Institute of Science, Rehovot 76100, Israel

¹⁰Department of Biology, Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02140, USA

¹¹Co-senior author

¹²Present address: Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

*Correspondence: aregev@broadinstitute.org

<http://dx.doi.org/10.1016/j.cell.2014.11.015>

SUMMARY

Cells control dynamic transitions in transcript levels by regulating transcription, processing, and/or degradation through an integrated regulatory strategy. Here, we combine RNA metabolic labeling, rRNA-depleted RNA-seq, and DRiLL, a novel computational framework, to quantify the level; editing sites; and transcription, processing, and degradation rates of each transcript at a splice junction resolution during the LPS response of mouse dendritic cells. Four key regulatory strategies, dominated by RNA transcription changes, generate most temporal gene expression patterns. Noncanonical strategies that also employ dynamic posttranscriptional regulation control only a minority of genes, but provide unique signal processing features. We validate Tristetraprolin (TTP) as a major regulator of RNA degradation in one noncanonical strategy. Applying DRiLL to the regulation of noncoding RNAs and to zebrafish embryogenesis demonstrates its broad utility. Our study provides a new quantitative approach to discover transcriptional and posttranscriptional events that control dynamic changes in transcript levels using RNA sequencing data.

INTRODUCTION

Dynamic changes in transcript levels are tightly regulated by the interplay of RNA transcription, processing, and degradation. Cells can produce complex dynamic mRNA patterns by changing one or more of these rates (Figure 1A). For example, either

increasing transcription or decreasing splicing or degradation rates can yield a similar temporal mRNA profile (Figure 1A, red). Compensatory changes in two (or more) of these rates can also leave the mRNA levels unchanged and thus diminish or obscure regulatory transitions, say if decreased processing counteracts increased transcription (Figure 1B). However, most studies only measure mRNA levels and tacitly focus on transcriptional regulation, excluding changes in RNA degradation or processing from consideration.

The many possible regulatory strategies raise important questions. How does each regulatory strategy combine a transcript's transcription, processing, and degradation rates to generate its expression pattern? Are genes with similar temporal mRNA profiles controlled by the same strategy? If not, what function do different strategies serve if their outcome (mRNA profile) is seemingly the same? Does local variation in transcription or splicing rates along a transcript's length regulate its expression? These questions are not fully understood, even for specific transcripts.

Technical and computational challenges have limited the availability of genome-wide dynamic data on RNA transcription, processing, and degradation. Methods for measuring RNA regulation rates in vivo typically require severe manipulations (Audibert et al., 2002; Core et al., 2008; Shalem et al., 2008; Singh and Padgett, 2009), impacting physiological relevance. Fractionation-based methods (Churchman and Weissman, 2011; Pandya-Jones et al., 2013) may be impacted by nonspecific RNA binding and coprecipitating proteins. Recently, several studies (Dölken et al., 2008; Eser et al., 2014; Rabani et al., 2011; Windhager et al., 2012) used short pulses of 4-thiouridine (4sU) RNA labeling to isolate newly-transcribed RNA and determine RNA kinetics during dynamic responses. Although most focused on RNA transcription and degradation, this strategy was also applied for RNA processing (Rabani et al., 2011; Windhager

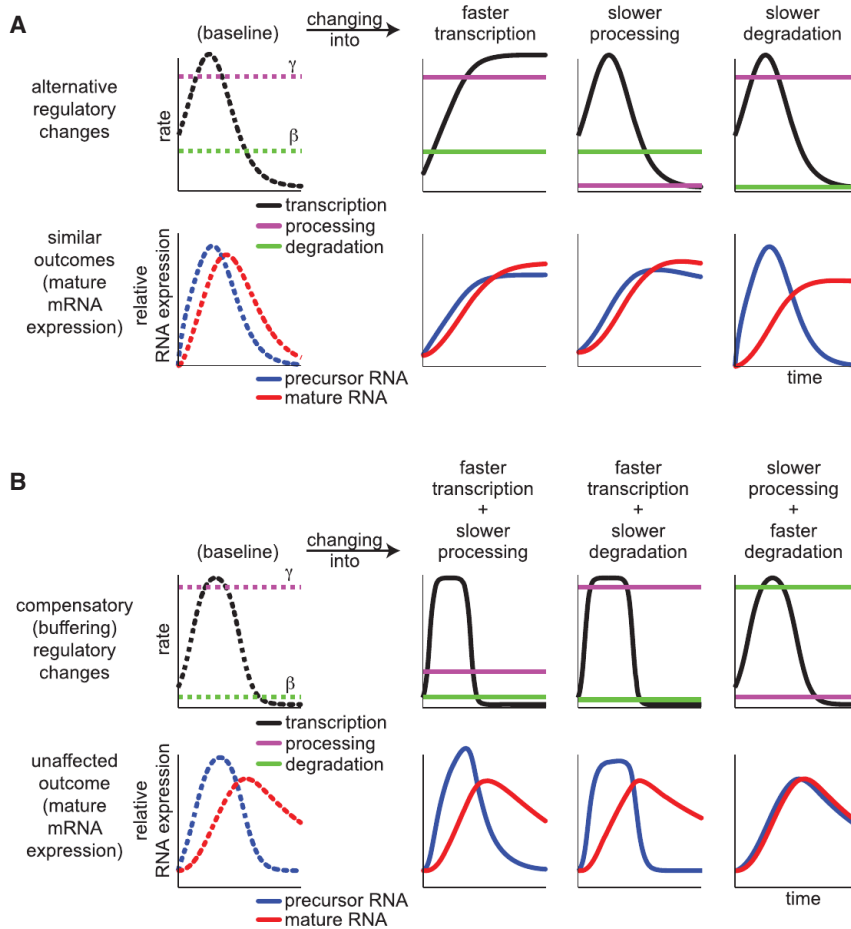


Figure 1. Dynamic Transitions in Mature RNA Levels Can Arise from Changes in Transcription, Processing, or Degradation

(A) Different regulatory changes can lead to a similar mRNA temporal expression profile. Top: transcription (black, RNA/min), processing (magenta, 1/min) and degradation rates (green, 1/min). Bottom: precursor (blue) and mature (red) RNA expression levels. Left (dashed lines): baseline reference expression. Three columns (solid lines): changes in each of three possible rates, lead to the same new, mRNA profile (solid red, bottom).

(B) Compensatory changes in two of three rates (rows as in A) leave mRNA levels (red, bottom) unchanged. Left column (dashed lines): reference expression; three columns (solid lines): changes from reference in two of three possible rates; mRNA levels (red, bottom) do not change versus baseline.

unique signal processing features. Finally, we apply DRILL to the early zebrafish transcriptome and to the regulation of unstable noncoding RNAs, establishing its general utility.

RESULTS

A High-Resolution Map of the Temporal Response of Mouse DCs to LPS

To monitor the relative regulatory contributions of RNA transcription, processing, and degradation, we sampled RNA from mouse DCs every 15 min, for the first

et al., 2012), albeit not dynamically. Moreover, although the excision rate of particular introns was described (Audibert et al., 2002; Singh and Padgett, 2009), there is no large-scale data on intron-specific processing rates, and few studies that measured RNA processing intermediates (Pandya-Jones et al., 2013; Rabani et al., 2011; Windhager et al., 2012; Zeisel et al., 2011) had insufficient resolution to study individual introns. Consequently, RNA-Seq analysis tools (Katz et al., 2010; Trapnell et al., 2009; Wang et al., 2008) are optimized for mature transcripts, but not unstable precursors.

Here, we generate a high-resolution map of the transcriptome in response to lipopolysaccharide (LPS) stimulation in mouse immune dendritic cells (DCs). We combine high-resolution sequencing of rRNA-depleted and of metabolically labeled RNA and a novel computational modeling approach (DRILL) to quantify (1) precursor and mature RNA levels at a splice junction resolution from rRNA-depleted sequencing counts, (2) kinetic rates of RNA transcription, processing, and degradation from metabolic labeling data, and (3) reliable RNA editing sites by detecting local differences in base composition between recently transcribed and overall RNA. Four regulatory strategies generate most (65%) expression patterns through changes in RNA transcription; noncanonical strategies with a dynamic posttranscriptional component affect a minority (35%) of genes and provide

3 hr of their response to LPS (Figure 2A; Experimental Procedures), following a short (10 min) metabolic labeling pulse with 4sU preceding the sampled time point. We isolated RNA from each sample in two ways: (1) RNA depleted of rRNA (RNA-Total) to measure total RNA regardless of its transcription time, and (2) 4sU-labeled RNA (RNA-4sU) that captures primarily RNA transcribed during the 10 min labeling pulse and is thus enriched for short-lived transcripts, including mRNA precursors and processing intermediates. We deeply sequenced each sample (80–200 million paired-end 101 base reads per sample) (Experimental Procedures; Table S1 available online). Although any time point is measured only once, we analyzed them jointly to minimize biases in any one sample.

A Model-Based Approach Quantifies the Abundance and Kinetics of Precursor and Mature Transcripts at Single Junction Resolution

We developed dynamic RNA life cycle (DRILL), a novel computational scheme to quantify transcript abundance and kinetic rates at the level of individual splice junctions in precursor and mature transcripts (Figures 2B, 2C, S1A, and S1B). DRILL consists of two consecutive modules.

First, a binomial model (Figure 2B; Experimental Procedures; Extended Experimental Procedures) uses RNA sequencing

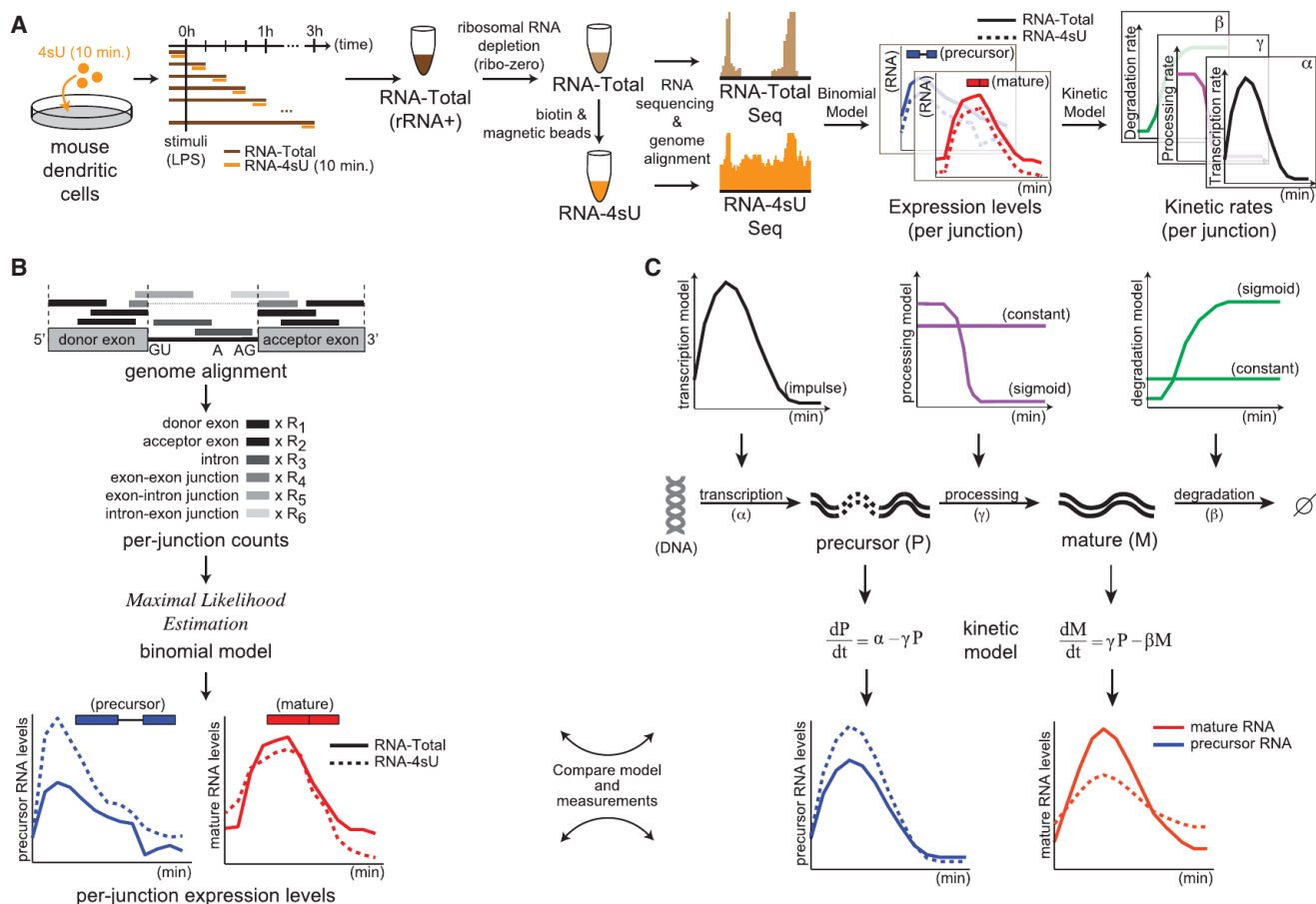


Figure 2. DRiLL Infers the Abundance and Kinetics of Precursor and Mature Transcripts at Single Junction Resolution

(A) A high-resolution map of the temporal LPS response. Orange: 4sU pulse and 4sU-RNA. Dark brown: sampled RNA; light brown: rRNA-depleted Total RNA; blue, red: inferred precursor and mature levels, respectively; black, purple, green: estimated rates of RNA transcription, processing, and degradation, respectively.

(B) Binomial model. Counts of sequencing reads that are located on exons, introns or the junctions between them (grayscale, dark to light) are used to infer, for each splicing junction, the abundance of transcripts with an unspliced precursor (P , blue) and mature junction (M , red), in either RNA-Total (solid) or RNA-4sU (dashed) samples.

(C) Kinetic model. Transcription makes a precursor (P , blue) of the junction (at some temporally changing rate α , black), and that product (P) is processed (at rate γ , purple, constant or temporally changing) into a mature transcript (M , red). Degradation (at rate β , green, constant or temporally changing) eliminates the mature (M) junction. Comparing the kinetic model estimates of P and M to their levels as inferred by the binomial model (red and blue, respectively), the model fits the kinetic parameters of a junction.

See also Figures S1 and S7 and Table S1.

(RNA-seq) counts to infer, for each splicing junction, the abundance of transcripts with an unspliced junction (precursor transcripts, P) and those with a fully spliced junction (mature transcripts, M), and, when appropriate, distinguish the relative abundance of several mature isoforms (M_1, M_2, \dots, M_k) that arise from a single precursor. Inference relies on separating the different sequencing reads that span an annotated junction by their location on exons, introns or the junctions between them. It applies independently to each RNA-seq sample and thus is applicable to any deeply sequenced RNA, but is most appropriate for rRNA-depleted samples (see Discussion).

Second, a dynamic model uses the estimated abundance of the precursor and mature junctions from different RNA populations to infer each transcript's kinetic parameters: transcription,

splicing, and degradation rates (Figure 2C; Experimental Procedures; Extended Experimental Procedures). In this model, transcription (α) produces a primary precursor of the junction that is subsequently processed by splicing into a mature junction and ultimately degraded. A precursor processing rate (γ) represents the junction's half-life at its unprocessed form, and a mature degradation rate (β) models the mature junction's half-life, balancing RNA processing and decay. While degradation is expected to be uniform across a transcript, because we model each junction separately, its "degradation rate" reflects a local stability that is affected both by its own maturation and by the mature transcript's decay. For example, if one junction within a transcript is spliced much faster than others, its "degradation rate" is lower than that of other junctions, simply because it

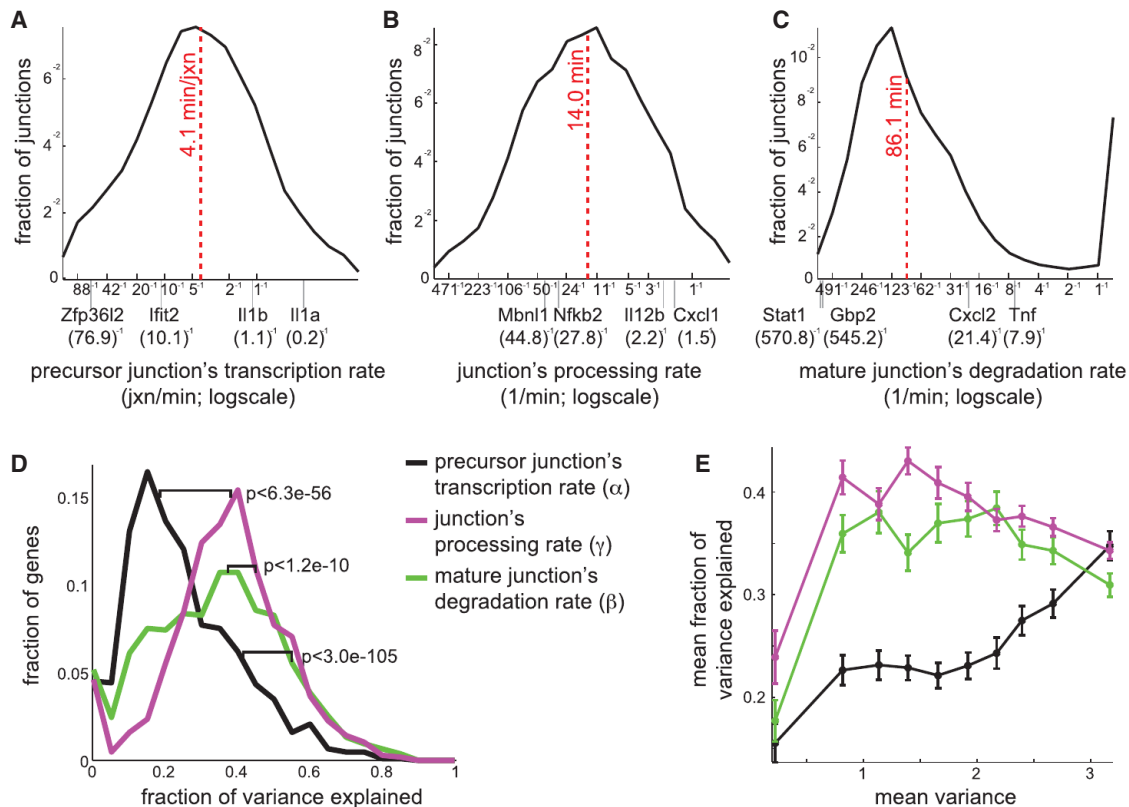


Figure 3. Genome-wide Kinetic Rates at Per-Junction Resolution

(A–C) Distribution of junction kinetic rates (x axis, log scale) predicted for 10,351 substantially expressed junctions (fraction of junctions, y axis). Example transcripts and half-life values in minutes are marked. Dashed line: median. (A) Precursor junction's transcription rates (jxn/min, x axis, log scale). (B) Junction's processing rate (1/min, x axis, log scale). (C) Mature junction's degradation rate (1/min, x axis, log scale).

(D) Distribution of the fraction of the variance between a gene's junctions that is explained by differences in transcription (black), processing (purple), or degradation (green) rates, in 1,693 genes with ≥ 2 junctions (fraction, y axis); p values: KS test.

(E) The mean fraction (y axis) of the variance between a gene's junctions that is explained by differences in its transcription (black), processing (purple), or degradation (green) rates, estimated in each of ten quantiles of genes (x axis) partitioned by mean variance between junctions. Error bars represent SE.

See also Figures S2 and S3.

starts its life as a “mature junction” earlier, while the rest of the transcript is still being processed (see below). Reliable estimation of kinetic parameters usually requires high-resolution temporal data of total and metabolically labeled RNA, but less data suffices under certain conditions (see below).

Large-Scale Analysis of RNA Regulatory Rates at Per-Junction Resolution

We first used DRiLL to quantify junction-specific kinetic rates for transcription, processing, and degradation of the top 10% of expressed genes with highest coverage (1,128 genes, encompassing 9% of all annotated junctions) (Figures S2A and S2B; Experimental Procedures) and hence can be analyzed most precisely (98% have tight confidence intervals).

DRiLL's inferred expression levels and rates were reproducible and accurate by several tests (Figure S1; Experimental Procedures). The transcription rates of an unprocessed junction range from tens of seconds to tens of minutes per junction (median of 4.1 min/junction; Figure 3A), well in line with recent measurements of RNA polymerase elongation rates in human HeLa

cells (Fuchs et al., 2014). The half-life of a precursor junction, reflecting its splicing rate, ranges from fractions of minutes to an hour (median of 14.0 min; Figure 3B), and agrees with few measured individual intron splicing rates in human (Singh and Padgett, 2009) and mouse (Audibert et al., 2002). The mature junction's half-life, reflecting the stability of the processed junction, ranges more widely from a few minutes to a few hours (86.1 min median; Figure 3C), and is typically longer than the precursor junction's half-life ($p < 1.7 \times 10^{-33}$, Kolmogorov-Smirnov [KS] test; median difference of 55 min; Figure S2C). Significant dynamic changes in processing and/or degradation rates are evident in 15% of junctions and are also faster on average (Figures S2D–S2F).

Differential Processing Efficiency Is a Major Source of Intratranscript Variation

Junctions from the same transcript are generally regulated jointly and thus have highly similar levels and associated rates compared to those on separate transcripts ($p < 8 \times 10^{-111}$, KS test), especially when comparing adjacent junctions ($p < 5 \times 10^{-250}$,

KS test; Figure S2H). Indeed, internal transcript differences account for only 40% of overall variation (Figure S2G) and contribute less to the total variance in our data than differences between transcripts.

Nevertheless, local events can give rise to variation in the level of different splicing junctions within a transcript. Transcriptional pausing or length differences can lead to changes in transcription rates between junctions (e.g., compare jxn 1 and 6 in *Tcfec*; Figure S3A), while differences in splicing efficiency between junctions would result in different half-lives of their individual precursor (e.g., jxn 1 and 3 in *Cxcl2*; Figure S3B) or mature (e.g., jxn 9 and 10 in *Zc3hav1*; Figure S3C) forms, with some junctions spliced long before the rest of the transcript matures (e.g., jxn 3 in *Il12b*; Figure S3D).

Globally, half-life differences explain most (75%) of the internal variation between junctions of the same transcript (37% by precursor and 38% by mature junction's half-life differences; Figure 3D), supporting differential splicing efficiency as a main source of internal transcript variability. Local differences in transcription explain only 25% of the variation, but their contribution is more prominent for transcripts with high internal variation (35% in the most variable transcripts quantile; Figure 3E). Indeed, correlation between individual junctions' and whole-transcripts' rates (see below) is higher when comparing transcription rates ($r = 0.6$) than for processing ($r = 0.38$) or degradation ($r = 0.47$) rates (Figure S4A).

Gene-Specific Regulatory Rates in the Dynamic RNA Life Cycle

We systematically studied dynamic RNA regulation in top 70% of annotated transcripts, (7,872 transcripts, spanning 76% of all junctions; Experimental Procedures) with at least a minimal coverage of their exons, introns, and junctions. Given the high similarity in expression of junctions within most transcripts, we took the median abundance across all junctions in a transcript as representative of the gene's dynamics and estimated from that the transcript's overall kinetics.

Faster (top 20%–30%) transcription, processing and degradation rates are typically associated with shorter transcripts with fewer (seven on average) and shorter (29 kb overall length on average) exons and introns, while slower rates (low 20%–30%) are associated with longer transcripts with a larger number (13 on average) of longer (54 kb overall length on average) exons and introns (Table S2). Very slow processing rates (top 10%) are associated with alternatively spliced transcripts, but unexpectedly also with short transcripts and short introns. Finally, transcription and processing rates are more highly correlated to each other (Figures S4B and S4C; $r = 0.61$, $p < 1 \times 10^{-40}$) than either is with degradation rates ($r = 0.48$ and 0.47 respectively; $p < 1 \times 10^{-40}$), consistent with a coherent regulatory coordination between the two biosynthesis steps (transcription and processing).

Most Genes Are Regulated by Transcription-Dominated Canonical Strategies

To understand how the regulatory steps are coordinated, we examined which regulatory strategies (Figure 1A) are predominantly used in the DC response. We clustered the genes into

22 groups based on their kinetic parameters (Figure 4A). The genes in each group use rates in a similar way to shape the dynamics of their final mature product (Figures 4A and S5A) and therefore share the same “regulatory strategy.” There are four temporal categories of mRNA profiles: transiently induced (groups 1–4), upregulated (groups 5–11), transiently repressed (groups 12–15), and downregulated (groups 16–21). More than half the genes in each category (Figure 4B) employ a single strategy, thus most patterns (65%) arise from just one of four strategies.

All four predominant regulatory strategies (Figure 4C) combine dynamic changes in transcription with temporally-constant processing and degradation rates (these rates change in only 4.3% and 10% of genes respectively, below). Genes with transiently induced mature RNA are enriched for inflammatory signaling proteins (e.g., *Tnf*) and transcription factors (e.g., *Nfkb*) and typically (group 3; 70%) arise from transient increases in transcription rates, combined with fast (constant) processing rates (e.g., *Ifrd*; Figures 4D and S5B–S5E). Upregulated genes are enriched for viral and interferon response genes and typically (groups 9–10, 62%) arise from an increase in transcription rate combined with constantly fast processing and constantly slow degradation rates (e.g., *Cpeb4*; Figures 4D and S5B–S5E). The transcriptional increase is commonly (44%) a “production overshoot” as previously reported in macrophages (Zeisel et al., 2011): strong transcriptional induction (transcription fold-change is at least twice as high as that of mRNA levels) that contributes to a fast accumulation of the mature transcript and can either be transient (clusters 5 and 8) or persistent (clusters 9 and 10) in our time scope. Transiently repressed genes are generally enriched for housekeeping genes, with canonical group 14 (71%) also specifically enriched for mitochondrial and vesicular genes. All are canonically regulated at two stages (e.g., *Atp6v*; Figure 4D): initially, there is little to no new transcription, and fast degradation eliminates preexisting mRNAs; subsequently, transcription increases rapidly and mRNAs accumulate again, albeit with a temporal delay. This expression profile replaces old transcripts with new ones, rather than accumulating on top of them. Finally, downregulated genes are enriched for proliferation and cell-cycle factors (e.g., EGFR signaling) and generally (groups 19 and 21; 53%) arise from a decrease in transcription rates combined with (constant) slow processing and fast degradation rates (e.g., *Coro1a*; Figures 4D and S5B–S5E). Slow processing rates either delay the effect of increases in transcription on final mRNA levels (in transiently repressed cluster 13), or buffer them such that they do not manifest in mRNA levels during our temporal span (in downregulated clusters 16, 18, and 20).

Alternative Regulatory Strategies Use Dynamic Regulation at Multiple Steps to Generate Similar Expression Patterns with Unique Functionalities

A minority of genes (35%) follow different regulatory strategies that often involve a dynamically regulated posttranscriptional component but seemingly result in the same mRNA patterns as canonical strategies. Regulation of mRNA through both transcription and RNA processing/degradation forms a feed-forward loop (FFL) (Figure 5A) that we use in simulation studies

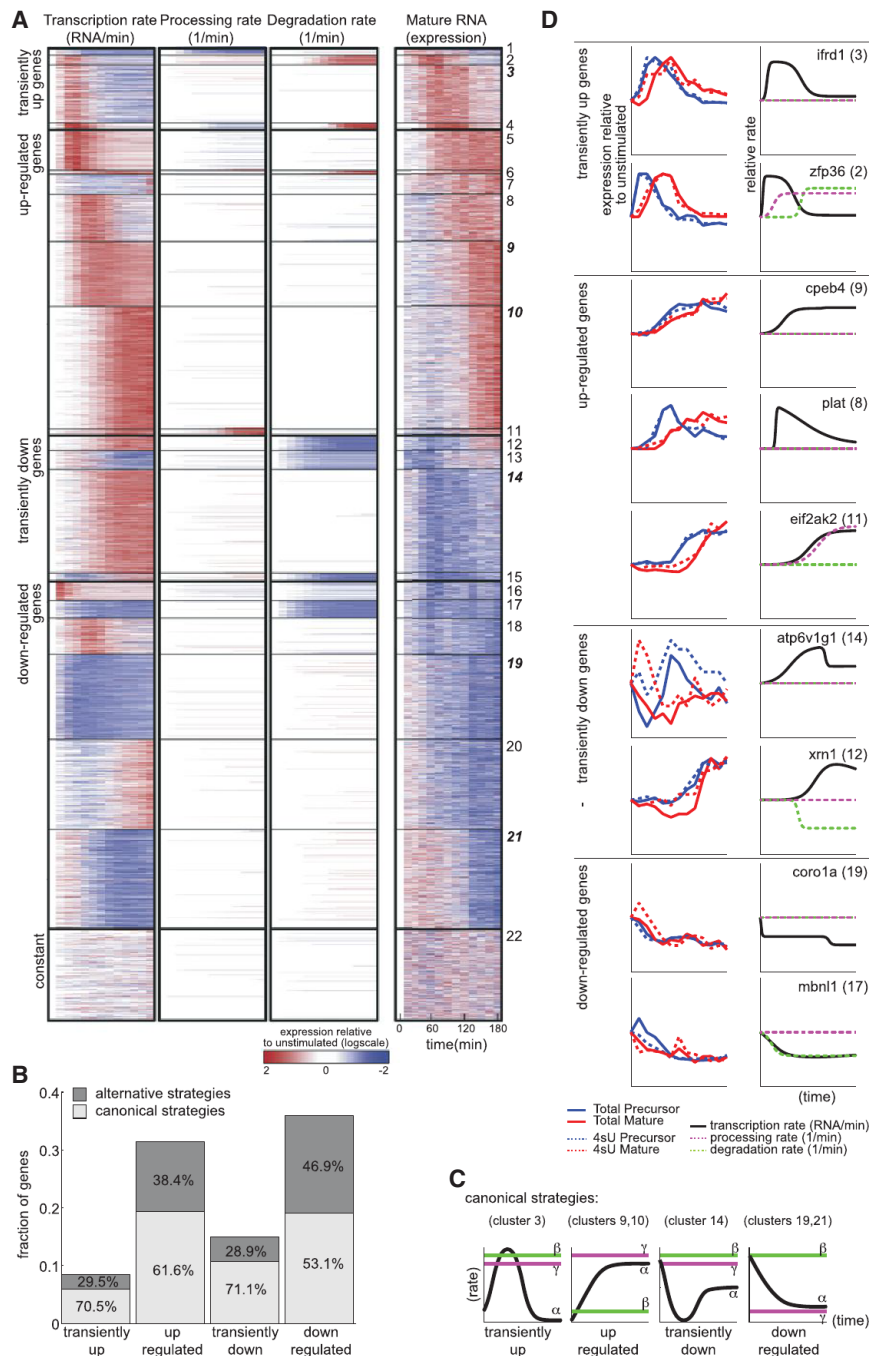


Figure 4. Regulatory Strategies that Generate Dynamic mRNA Profiles

(A) Dynamics of transcription (left), processing (middle), and degradation (right) kinetic rates predicted by the kinetic model and mRNA levels inferred by the binomial model, relative to unstimulated (t_0) control (white: t_0 , red: 2-fold above t_0 , blue: 2-fold below t_0 ; log scale), for each of 7,872 expressed genes (rows) during 3 hr of the response (columns). Genes are divided into 22 groups (solid black lines), in four modes of mRNA regulation (dashed black lines, from top to bottom): transiently up, upregulated, transiently down and downregulated.

(B) Fraction of genes (y axis) using canonical (light gray) or noncanonical (dark gray) strategies in each of the four modes (x axis). Fraction of genes within each mode is marked.

(C) Canonical regulatory strategies. Typical transcription (α , black), processing (γ , purple), and degradation (β , green) rates of canonical strategies in each of the four modes.

(D) Example genes (name on top, group in brackets) from canonical and non-canonical strategies. Right plots: t_0 -relative expression (y axis) of a gene's precursor (blue) and mature (red) RNA inferred by the binomial model for RNA-total (solid) and RNA-4sU (dashed). Left plots: kinetic parameters of a gene (relative to rate at t_0 , y axis): transcription (black), processing (dashed purple), and degradation (dashed green).

See also Figures S4, S5, and S6 and Tables S2 and S5.

due to an initially low processing rate, but once processing rate increases, mRNA peaks quickly and very highly. Next, degradation rate increases and leads to a quick removal of these transcripts. The canonical regulatory strategy in cluster 3 generates an apparently similar mRNA dynamics through regulation of transcription rates alone, while processing and degradation remain constant (e.g., *ifrd*; Figure 4D). However, important differences between those groups suggest a functional role for the alternative strategy. First, genes in cluster 2 have a much higher maximal expression than those in cluster 3 (3-fold higher median; Figures S5B–S5E), possibly due to a prolonged period of maximal transcription (Figure 5B). Furthermore, simulations suggest that a coupled increase in transcription, processing, and degradation rates maintains the same peak expression level even for noisy signals, whereas if only transcription is regulated, peak expression is much lower when the signal is noisy (Figure S5F). While both clusters 2 and 4 display a delayed increase in degradation following transcriptional induction, which produces sharp peaks of RNA levels as previously described (Rabani et al., 2011), our current analysis reveals

(Experimental Procedures) to compare with matching canonical strategies and determine the function of the posttranscriptional component.

Among transiently induced genes, those in cluster 2 (enriched for inflammatory response genes) quickly reach a maximal transcription rate and maintain it, while their mRNA levels increase and decrease more gradually, because of dynamic regulation of both processing and degradation rates (e.g., *Zfp36*; Figure 4D). As transcription increases, the precursor accumulates

longed period of maximal transcription (Figure 5B). Furthermore, simulations suggest that a coupled increase in transcription, processing, and degradation rates maintains the same peak expression level even for noisy signals, whereas if only transcription is regulated, peak expression is much lower when the signal is noisy (Figure S5F). While both clusters 2 and 4 display a delayed increase in degradation following transcriptional induction, which produces sharp peaks of RNA levels as previously described (Rabani et al., 2011), our current analysis reveals

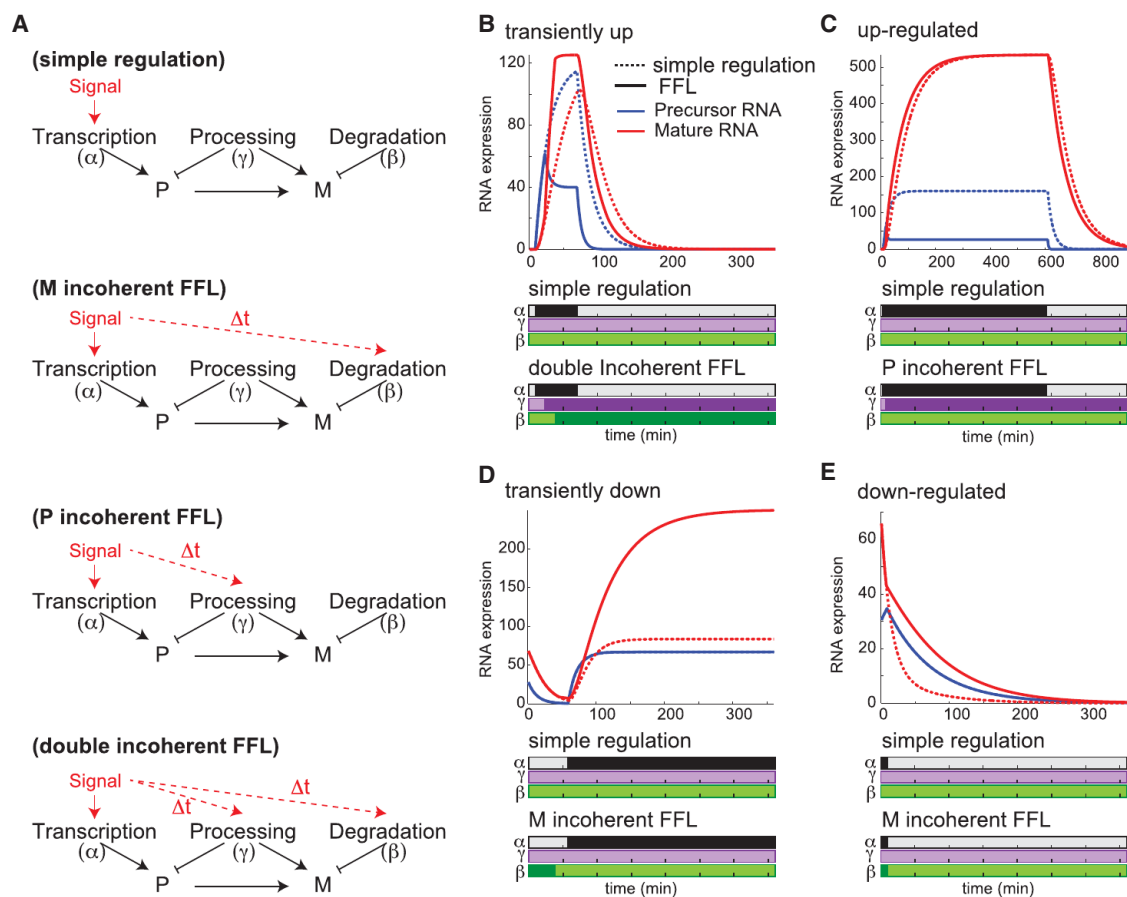


Figure 5. Simulations Suggest Functional Role of Alternative Regulatory Strategies

(A) Canonical and alternative strategies. Top to bottom: simple regulatory strategy where only transcription rates change dynamically (red arrow); incoherent feed-forward loop (FFL) regulation of mRNA expression with additional temporal changes in degradation rates (dashed red arrow, temporally delayed); incoherent FFL regulation of precursor expression with additional temporal changes in processing (dashed red arrow, temporally delayed); a double incoherent FFL with temporal changes in transcription, processing, and degradation rates.

(B–E) Comparing simple (dashed lines) and alternative (solid lines) strategies: a double incoherent FFL (B), a precursor incoherent FFL (C), and a mature RNA incoherent FFL (D and E). Top: temporal (x axis, minutes) precursor (blue) and mature (red) RNA expression (y axis) by either strategy. Bottom: temporal (x axis) transcription (black), processing (purple) and degradation (green) rate changes relative to unstimulated cells by a simple (top) or alternative (bottom) strategy. See also Figure S5.

that processing rates increase in cluster 2 but decrease in cluster 4 and contribute to the shutoff rather than the onset phase.

Among upregulated genes, transcription rates of cluster 11 genes steadily increase, but their processing rate is also dynamically regulated. At the beginning of the response, a low processing rate allows the unprocessed transcripts to accumulate (e.g., *Eif2ak2*; Figure 4D). Subsequent increase in both transcription and processing rates results in a faster accumulation of the mature mRNA to higher levels (2-fold on average; Figures S5B–S5E) and a faster predicted future shutoff (Figure 5C) than the canonical strategy of clusters 9 and 10. This regulatory strategy is more sensitive to expression noise (Figure S5G), which might explain why it is not implemented for the lower expressed, and thus noisier, genes in clusters 9 and 10. Another alternative strategy, in clusters 5 and 8, gives rise to a very similar mRNA pattern: instead of a steady increase in transcription, an early short burst of high transcription rates levels off to a moderate

rate (a transient “production overshoot,” e.g., *Plat*; Figure 4D). Coupled with constantly slow degradation rates, mRNAs that are relatively long lived accumulate quickly.

Transiently repressed genes in clusters 12, 13, and 15 have dynamically regulated transcription and degradation rates. Unlike the canonical strategy (cluster 14) where degradation rate is constantly high (e.g., *Atp6v*; Figure 4D), here degradation rate is only high initially, contributing to eliminating all existing transcripts and then slows down (e.g., *Xm1*; Figure 4D). The reduced degradation rate is combined with increased transcription rate and leads to RNA accumulation and eventually to a higher steady-state expression compared to the canonical strategy (2-fold on average; Figures S5D and S5B–S5E). Similarly, in the downregulated genes of clusters 16 and 17 both transcription and degradation rates decrease (e.g., *Mbn1*; Figure 4D), resulting in a slower decrease of mRNA levels and higher steady-state mRNA levels (Figure 5E) compared to the canonical

strategy (clusters 19 and 21). Cluster 16 is enriched for many housekeeping genes, which cells must maintain even if at lower levels. This strategy lowers the energetic price of expression, but at the cost of slower regulation.

Predicting Molecular Regulatory Mechanisms by Integration of RNA and Protein Life Cycle Data

To explore some of the molecular mechanisms governing distinct regulatory strategies, we analyzed the clusters (Figure 4A) for their correlation to changes in putative regulatory proteins with a known RNA binding activity as measured by pulsed SILAC proteomics (M.J., M. Rooney, N.H., and A.R., unpublished data). Seven RNA binding proteins were each highly correlated ($r > 0.98$) to changes in RNA degradation or processing rates of at least ten transcripts.

Several lines of evidence support a role for one of these candidates, Tristetraprolin (TTP, *Zfp36*) in regulating changes in RNA degradation rate in DCs. First, TTP is a known regulator of RNA stability (Brooks and Blackshear, 2013) of several key immune genes (Lai et al., 2006) and responds to many LPS-activated signaling pathways (Lai et al., 2006). Second, the consensus ARE heptamer (UAUUUUAU) associated with RNA destabilization by TTP (Lai et al., 2006) is present in 3'UTRs of 58/109 genes ($p < 2.3 \times 10^{-13}$) with a predicted increase in degradation rates. Finally, consistent with TTP's known autoregulatory role (Brooks and Blackshear, 2013), TTP's RNA degradation rate increases in the LPS response (Figure 4D), as does that of its most well-established target, *Tnf* (Carballo et al., 1998) (data not shown).

TTP Is Required for Upregulation of Degradation Rates in Transiently Induced Genes

To test the hypothesis that TTP regulates RNA degradation rates during the LPS response, we measured RNA levels and transcription rates in DCs derived from either normal (wild-type [WT]) or from homozygous TTP knockout mice (TTP-KO) every 15 min along a 3 hr time course of their response to LPS (Figure 6A). We used the nCounter to measure each of 267 signature immune genes (Experimental Procedures). Transcripts regulated by TTP should demonstrate a changed degradation rate between WT and TTP-KO cells. To identify and quantify these changes, we used a novel molecular model of a *trans* regulator of mRNA degradation (here, TTP) (Figure 6B; Experimental Procedures; Extended Experimental Procedures).

TTP is predicted by the model to regulate the degradation of 36 transcripts within our "signature set": in WT cells, degradation of these transcripts increases at 60–90 min poststimulation, but in TTP-KO cells they have only minimal changes of degradation (Figure 6C). These include 7/11 known TTP targets (Brooks and Blackshear, 2013) ($p < 10^{-4}$) and are enriched (17/36 targets, $p < 5.3 \times 10^{-11}$) with upregulated degradation rates from RNA-seq data (above). Furthermore, our model's estimated regulator activity function agrees with measured (M.J., M. Rooney, N.H., and A.R., unpublished data) changes in TTP protein levels (Figure 6D), our predicted K_m values negatively correlate (Spearman $\rho = -0.21$, $p < 7.7 \times 10^{-4}$) with known TTP binding preferences (Brooks and Blackshear, 2013), and our estimated Hill coefficient ($n = 2.9$) that suggests that TTP binds cooperatively, is consistent with previous studies (Brooks and Blackshear, 2013) and with

enrichment of targets for multiple occurrences of the TTP binding pentamer (14/36 targets, $p < 3.0 \times 10^{-5}$). Our analysis suggests that TTP may also independently affect its targets' transcription, because transcription rates of the 36 predicted TTP targets significantly decrease in TTP-KO cells (3-fold versus 1.3-fold on average for nontargets; Figure 6E). This is likely an indirect effect, at either the transcription or processing level.

Revealing Reliable RNA Editing Sites in Noncoding Portions of LPS Response Transcripts

We used our data to identify other steps in the RNA life cycle, such as RNA editing events, whose detection by high throughput sequencing (Danecek et al., 2012; Li et al., 2009, 2011; Neeman et al., 2006) raised substantial debates (Kleinman and Majewski, 2012; Lin et al., 2012; Pickrell et al., 2012) due to difficulties to computationally control for the current error rates in RNA-seq. As an alternative (Figure 7A; Experimental Procedures; Extended Experimental Procedures), rather than comparing to DNA sequences, we compared base changes between two RNA-seq experiments: RNA-4sU-seq and RNA-total-seq, expecting that editing changes will be more prominent in RNA-total than in newly transcribed RNA-4sU, but that error-prone positions will be equally affected in both samples.

We found 70 editing sites in 43 loci across the DCs transcriptome (Table S3), a substantially lower number than estimates in human (Li et al., 2009), and supported them by several lines of evidence. First, a lower editing level is expected in mouse in the absence of primate-specific Alu repeats (Neeman et al., 2006). Second, as an internal positive control, nucleotide changes called in newly transcribed RNA-4sU are almost exclusively (315/319) C to T modifications that are known (Hafner et al., 2010) to arise when sequencing 4sU residues. Conversely, predicted edits (Figure 7B) are mostly (61/70) known deaminations: either A to I (38 A/G changes and 11 complement T/C changes, which likely arise from sequencing strand biases), or C to U (six C/T changes and six complement G/A changes). Surrounding sequences are enriched for forming stem-loop structures with an upstream sequence ($p < 5.7 \times 10^{-12}$), but not with a downstream sequence, consistent with the known binding preference of adenosine deaminase (ADAR). Third, none of the edited sites affects an annotated protein sequence (Figure 7B), while many sites (17/43) are associated with annotated and putative pseudogenes (e.g., *Taldo1*, *Psme2b*), which often contain multiple edited positions (8/17); this is consistent with a postulation that editing controls the expression of many transposable elements in human (Neeman et al., 2006). Finally, mass spectrometry detected 18 peptides that match a reading frame within one of the putative pseudogenes (within the intron of the gene *Ccrn4l*) and confirmed a predicted G/U editing that changes a Valine residue into a Leucine. All other sites are located in non-protein-coding portions of expressed genes (22/43 in 3'UTRs, 4/43 in introns) and potentially contribute to their posttranscriptional regulation.

Applying DRILL to Diverse Systems and Noncoding Transcripts Demonstrates Its General Utility

To demonstrate DRILL's wide applicability, we used it to examine the regulation of noncoding RNAs in our system and

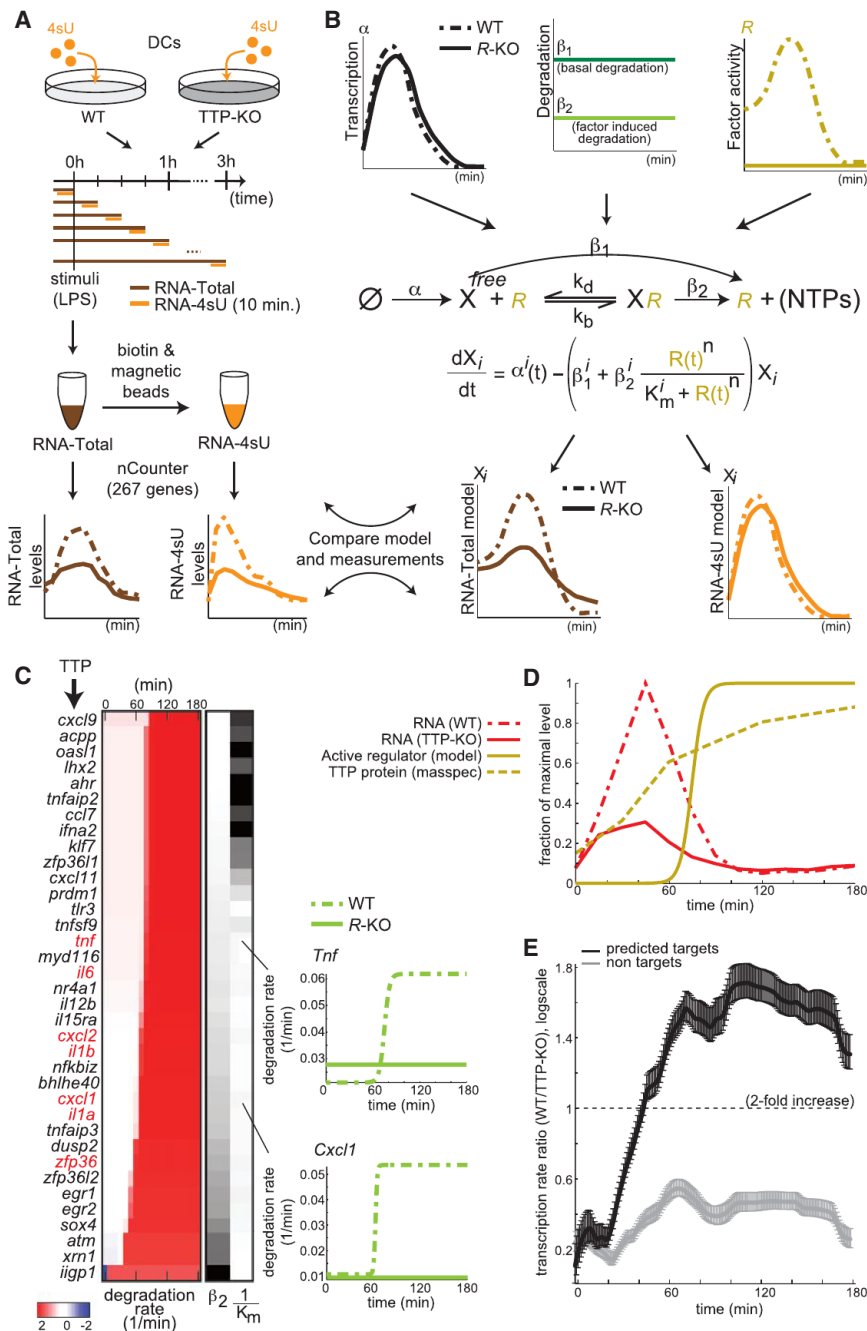


Figure 6. TTP as a Regulator of Dynamic RNA Degradation Rates

(A) Method overview. 4sU (orange) and total (brown) RNA are sampled from DCs derived from wild-type (WT, light gray) and TTP-KO (dark gray) mice, following LPS stimulation and short (10 min) metabolic labeling pulses, and quantified for a 267 transcript signature by the nCounter.

(B) Kinetic model of factor-induced RNA degradation. Gene X is transcribed at rate α (black) that differs in WT (dotted) or TTP-KO (solid) cells and is degraded either at basal rate β_1 (dark green) from the unbound state (X^{free}), or through factor-mediated (R , yellow, commonly an RBP) degradation (rate β_2 , light green) from the bound state (XR), in either WT (dotted) or TTP-KO (solid, inactive) cells. The regulator's association and dissociation constants (k_b , k_d) determine the binding efficiency (K_m). We optimize the parameters per gene by comparing the model predictions (bottom, RNA-Total: brown; RNA-4sU: orange) to the nCounter measurements.

(C) Thirty-six predicted TTP targets. Rows: genes (left; red: known TTP targets). Left heatmap: estimated WT degradation profiles (relative rate; red: high; blue: low) at 13 time points (columns). Right heatmap: predicted $1/K_m$ (binding affinity, left column) and β_2 (factor-induced degradation, right column).

(D) Predicted levels of the active regulator protein (solid yellow), TTP protein levels measured in WT cells (dashed yellow; average of two replicates), and TTP RNA levels in WT (dashed red) and TTP-KO (solid red) cells.

(E) Mean ratio of predicted transcription rate (WT versus TTP-KO rate; y axis; log scale) over time (x axis) for 36 predicted TTP targets (black) and nontargets (gray). Error bars represent SE. See also Table S4.

data, we found that eRNAs are transcribed at a very high rate, but are also very quickly degraded. Conversely, lincRNAs are transcribed and processed at comparable rates to protein coding genes, but are significantly less stable (Figures S6C–S6I). This could help explain how lincRNAs are both lowly expressed and tissue-specific (Cabili et al., 2011).

of maternally deposited versus zygotic transcripts in early embryogenesis.

First, we used DRILL to dissect the regulation of unstable noncoding regulatory RNAs in DCs (Figure S6). Although both enhancer-associated RNAs (eRNAs) (Kaikkonen et al., 2013) (Figure S6A) and long intergenic noncoding RNAs (lincRNAs) (Carpenter et al., 2013) (Figure S6B) were implicated in key innate immune functions, including in DCs, their regulation has not been extensively studied in any system, because they are usually lowly expressed (Carninci et al., 2005). Applying DRILL to these newly annotated noncoding transcripts in our

Second, we used DRILL to analyze transcriptome dynamics during early zebrafish embryogenesis (Figure S7). Embryos initially rely on maternally provided mRNAs and only activate zygotic (embryonic) transcription ~ 3 hr postfertilization (hpf) (Lee et al., 2014; Schier, 2007). Using rRNA-depleted RNA-seq data (Lee et al., 2013), DRILL distinguished maternal from zygotic mRNAs (Figures S7A and S7B); using polyA⁺ RNA-seq (Pauli et al., 2012; Zhang et al., 2014), DRILL estimated the onset time and rate of decay of maternally provided messages (Figures S7C–S7E). We find two major waves of degradation of maternal messages: immediately after fertilization (0–1 hpf, 18%) or after

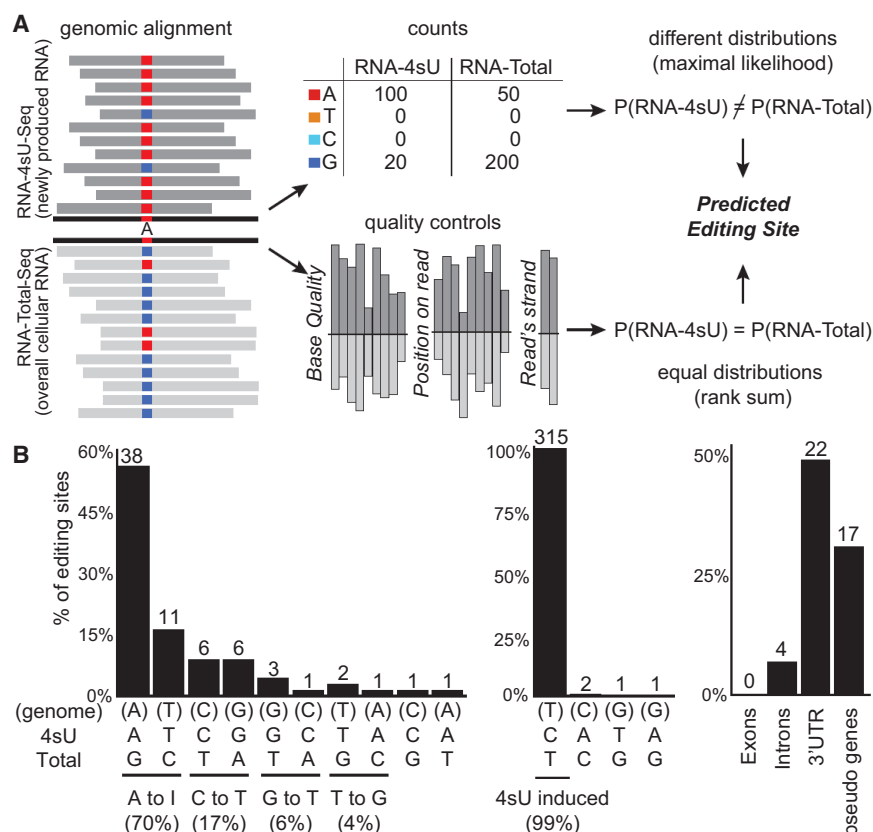


Figure 7. High-Resolution Metabolic Labeling Can Reliably Detect RNA Editing

(A) Method for detecting editing sites. We search for positions where the distribution of sequenced nucleotides is different in RNA-4sU-seq (dark gray, top) and RNA-total-seq (light gray, bottom) using maximum likelihood estimation (top row) and also require that other measures associated with base quality distribute evenly between the two samples (bottom row).

(B) Distribution of predicted editing sites (% of sites, y axis). Left: nucleotide changes in RNA-total (editing sites, nucleotide changes on x axis; top: genomic base, middle: RNA-4sU base, bottom: RNA-total base), middle: nucleotide changes in RNA-4sU data (4sU induced base changes). Right: distinct annotations associated with RNA-total nucleotide changes. Number of sites is marked. See also Table S3.

the maternal-to-zygotic (MZT) transition (3–5 hpf, 47%). Post-MZT decaying mRNAs are degraded faster than early decaying mRNAs (KS test $p < 10^{-29}$; Figure S7E) and are selectively enriched in their 3'UTR for seed sequences for miR-430 ($p < 8.3 \times 10^{-11}$), a microRNA involved in the degradation of maternal mRNAs (Giraldez et al., 2006). This suggests that different degradation pathways are active before versus after MZT. Indeed, early (2–4 hpf) polyA tail lengths of maternal mRNAs (as measured in Subtelny et al., 2014) correlate to their ribosomal occupancy (as measured in Chew et al., 2013; Figure S7F), but later (4–6 hpf) lengths are correlated with mRNA stability (Figure S7G). These findings support and extend the idea (Subtelny et al., 2014) that zebrafish posttranscriptional mechanisms change from a maternally derived control over mRNA translation into a zygotic regulation of mRNA stability.

DISCUSSION

We present a novel approach (Figure 2) that combines high-resolution RNA labeling and sequencing with advanced computational modeling (DRiLL) and uses it to study the regulatory strategies that generate temporal RNA levels during the LPS response.

Quantitative Dissection of the RNA Life Cycle in Dynamic Responses

DRiLL uses RNA-seq data to predict the frequency of mature and alternative transcripts and of their unstable precursors and

processing intermediates, which are mostly disregarded in other transcriptome analysis tools (Katz et al., 2010; Trapnell et al., 2009; Wang et al., 2008). As rRNA depletion becomes increasingly popular, especially when RNA quality is low (Adiconis et al., 2013), DRiLL will help researchers to explore transcriptomes at unprecedented depth and resolution. When temporal metabolic labeling data is also available, DRiLL further predicts kinetic transcription, processing, and degradation rates, both between transcripts and within transcripts (per-junction). This can be extended to other aspects of the RNA life cycle, such as RNA editing. DRiLL is broadly applicable, as we demonstrated for unstable long noncoding RNA (lncRNA) and for maternally provided mRNAs in zebrafish embryogenesis. Our genomic portal (<http://www.broadinstitute.org/rnalifecycle>) provides the scientific community with ready access to our analysis and tools.

Although the levels quantified by DRiLL are reproducible and reliable by several tests, they can be impacted by noise and biases in sequencing data, variations in coverage along genes and considering paired-end reads as independent observations. Introns retention in mRNAs can lead to further inconsistencies between junctions. Simplifying assumptions of the kinetic model (e.g., that global RNA levels in cells remain constant upon LPS stimulation, or that individual junctions are independently regulated) would affect our estimated rates, but would not change the ranking between genes (as all estimates will be similarly affected by such global events). Using a likelihood ratio test to select between constant and dynamic rate models also reduces DRiLL's sensitivity to detect changes in lowly expressed genes.

Key Principles of Temporal RNA Regulation in Mammalian Cells

We determined the key regulatory strategies that DCs implement to generate their mRNA outputs (Figure 4), demonstrating how similar or correlated mRNA profiles are generated in distinct

ways and hypothesized on their possible distinct functional utility (Figure 5). Our extensive data set can be further combined with other genome-scale data in this system. For example, decreased degradation rates in downregulated cluster 16 counteract transcriptional repression and increase the new steady state levels. Pulsed SILAC measurements of Cluster 16 proteins (M.J., M. Rooney, N.H., and A.R., unpublished data) also show an increase in their translation rate upon LPS stimulation, possibly as a second level of posttranscriptional buffering of their transcriptional repression.

Our work provides new and effective quantitative tools to study RNA dynamics at both transcript and per junction resolution from RNA-seq data and generates a unique view of the different kinetic strategies that cells use to coordinate transcriptional and posttranscriptional events and regulate transcript levels during a dynamic response.

EXPERIMENTAL PROCEDURES

DCs Culture and Sample Collection

All animal protocols were reviewed and approved by the MIT/Whitehead Institute/Broad Institute Committee on Animal Care (CAC protocol 0609-058-12) and by the Institutional Animal Care and Use Committee of the National Institute of Environmental Health Sciences (ASP protocol 97-06) for WT and TTP-KO mice. DCs culture and treatment, RNA sample collection, and 4sU-labeled RNA isolation were done as described in Rabani et al. (2011) with the following modifications. We added 4sU to a 500 μ M final concentration for 10 min before RNA collection. For RNA-seq, 10 μ g total RNA from each sample was depleted of rRNA by RiboZero (Epicenter), a 100 ng aliquot was kept for sequencing, and 4sU purification was done for the remainder of the sample.

RNA-Seq, Read Mapping, and Annotation

RNA-seq libraries were constructed by dUTP second strand protocol (Levin et al., 2010), sequenced by Illumina HiSeq2000 with paired-end, 101 bp reads (Table S2). We align reads to the mouse reference genome (NCBI37/mm9) using TopHat (Trapnell et al., 2009) with default parameters. We use polyA+ RNA-seq data (Garber et al., 2012) to reconstruct mRNA annotations with Trinity (Grabherr et al., 2011) and Cufflinks (Trapnell et al., 2010) and collect all annotated mouse transcripts (Refseq and UCSC genes NCBI37/mm9) (Rhead et al., 2010) that matched a reconstructed transcript. The oblong-stage (3.7 hr postfertilization) zebrafish RNA-seq sample was prepared as described in Pauli et al. (2012) and Zhang et al. (2014).

nCounter Sample Preparation and Data Processing

nCounter sample preparation, capture, and analysis were done as described in Rabani et al. (2011), with the following modifications. Our code set (Table S4) detects 246 signature LPS transcripts (Amit et al., 2009) and 21 control genes with constant basal expression levels (nine of which used for normalization), via a probe that matches their exon sequence (captures their pre-mRNA and multiple mature mRNA isoforms). For 30/246 transcripts, we had a second probe that matches their intron sequence and captures their precursor.

Substantially Expressed Genes

We define a splicing junction as substantially expressed if all its exons, introns, and the junctions between them have normalized counts (reads per kilobase of transcript per million reads mapped [RPKM]) sums (of all times and all RNA-Total or RNA-4sU samples) above their respective thresholds (10% or 70% substantially expressed genes). We take all genes with at least one substantially expressed junction.

Precursor and Mature RNA Abundance

We count sequencing reads that span an annotated junction by their location on either exons, introns, or the junctions between them, and use these counts

to quantify, for each splicing junction, the abundance of transcripts with an unspliced junction (precursor) and those with a fully spliced junction (mature). We use a binomial model in which the frequency of precursor and mature RNA directly relates to the probability of observing a given number of reads at each location, considering the depth of the sequencing library and the genomic lengths. We use derivative-free methods ("Nelder-Mead simplex algorithm" as implemented in MATLAB) to find the expression levels that are most likely to generate these read counts. We extend this to annotated alternative splicing and predict the relative abundance of several mature isoforms that arise from a single precursor junction. We apply this to the dynamic sequencing data of all substantially expressed junctions, independently for each RNA-seq sample.

Quantifying Transcript Kinetics

Our kinetic system model describes the time evolution of a junction's precursor (P) and mature (M_1, \dots, M_n) mRNA by its transcription (α), degradation (β), and processing (γ) rates:

$$\begin{aligned}\frac{dP}{dt} &= \alpha(t) - \sum_i \gamma_i(t)P \\ \frac{dM_i}{dt} &= \gamma_i(t)P - \beta(t)M_i\end{aligned}$$

We use gradient descent optimization to find the model parameters ($\theta = [\alpha, \beta, \gamma, X_0]$) that minimize the difference between the kinetic model predictions of precursor and mature transcripts levels to their direct estimates from RNA-seq by the binomial model (above). We compare four alternative hypotheses in which rates are either constant or change over time through a likelihood ratio test to identify genes in which dynamic changes in one or both rates significantly ($p < 0.01$) contribute to temporal changes in overall RNA levels and assign them with a time-dependent, rather than a constant, rate function. We apply this model using all temporal total and 4sU RNA levels (26 samples) either per an entire transcript or per a specific junction.

Model Fit, Reproducibility, and Accuracy

We use a goodness of fit test (χ^2 test) with the null hypothesis that the data is governed by the estimated binomial model and find minimal discrepancy ($p < 0.01$) with sequencing counts in >70% of our data, with mostly tight confidence intervals. Spearman correlation to an independent biological replicate set of nCounter measurements confirms reproducibility ($r > 0.73$). As expected, shorter-lived precursors are enriched in RNA-4sU samples and mature junctions in RNA-Total and RNA-polyA samples. The rate predictions are robust to normal additive error (estimated from genome-wide data), with a tight fit ($p < 0.01$, χ^2 test) in >90% of junctions. Confidence intervals are tight (least square error) by bootstrapping for 15 representative examples. Model predictions fit well to two unseen test data sets: polyA+ RNA-seq and nCounter data, taken at times within and beyond the scope of our training set. Predicted rates are significantly correlated with earlier predictions (Rabani et al., 2011) (degradation: $r = 0.39$; processing: 0.23), despite different time scale, resolution, and modeling.

Functional Enrichments

We test enrichment using a hypergeometric p-value (for binary features) or the KS test (for numerical features) and a 5% false discovery rate (FDR) across all tested annotations or all "substantially expressed" genes, respectively (Table S5). We calculate functional enrichments of rates by splitting all rates (at all times and all genes) into ten quintiles, assigning the most abundant quintile (across times) per gene and using hypergeometric p value.

Clustering

We first cluster (k-means clustering as implemented in MATLAB) a subset of 17% highly expressed genes (1,305 genes) after standardizing its \log_2 (expression) and/or \log_2 (rate) temporal data. We iteratively increase the number of clusters as long as none of the clusters has <2% of the genes. We assign each of the other genes into the same cluster of the gene in the initial subset with which it has a maximal Pearson correlation.

Simulation Studies

We simulate expression data using our kinetic model and characteristic kinetic parameters of RNA transcription, processing, and degradation rates. All rate functions are modeled as step functions with a basal rate and an active rate that is used only when an external signal exists. Input signal is modeled by a binary (0/1) function, and noise is introduced by random changes to its values. We simulate a temporal delay of processing and degradation response to an external signal by switching to the active rate only when precursor RNA levels (for processing) or mature RNA levels (for degradation) exceed a predefined threshold.

Factor-Dependent RNA Degradation

We model factor-dependent regulation of RNA degradation by a nonlinear Hill function with two constants: basal degradation of the unbound transcript (β_1) and factor-mediated degradation (β_2) that also depends on the unbound regulator's concentration. We fit two alternative models to both the WT and KO measurements and use a likelihood ratio test ($p < 0.01$) to select (for each gene) between the null hypothesis of a constant, factor-independent, degradation rate ($\beta_2 = 0$ in WT and KO), and a dynamic, factor-dependent regulation ($\beta_2 > 0$ in WT only).

RNA Editing

We search for edited positions with a different distribution of sequenced nucleotides between RNA-total and RNA-4sU (maximum likelihood test), but an equal distribution of base quality (Wilcoxon rank sum test), location on read (Wilcoxon rank sum test) and strand assignment of read (Fischer exact test).

ACCESSION NUMBERS

The Gene Expression Omnibus accession number for the RNA-seq data and processed files reported in this paper is GSE56977.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.11.015>.

AUTHOR CONTRIBUTIONS

M. Rabani, I.A., N.F., and A.R. conceived and designed the study. M. Rabani conducted the experiments. M. Rabani, N.F., and A.R. designed the computational methods. M. Rabani developed and implemented the computational methods. R.R. made the cell cultures. M.J. and M. Rooney conducted and analyzed the mass-spec experiments, which were designed with A.R. and N.H. D.J.S. and P.B. contributed the TTP-KO mice and helped design and interpret the relevant experiments. N.H. contributed experimental methods and reagents. A.F.S. and A.P. supported the zebrafish analysis. A.P. prepared the zebrafish RNA-seq sample.

ACKNOWLEDGMENTS

We thank the Broad Genomics Platform for sequencing and C. de Boer for comments on the manuscript. Work was supported by William Asbjornsen Albert Memorial and Xerox-MIT fellowships (M.R.), the Swiss National Science Foundation (SNSF) for advanced researchers and Marie Skłodowska-Curie IOF fellowships (M.J.), NIH R01GM056211 (A.F.S.), the Human Frontier Science Program (HFSP) and NIH K99-HD076935 (A.P.), the Centers of Excellence in Genomic Sciences (CEGS) (1P50HG006193-01, N.H., I.A., A.R.), and Pioneer Award (DP1OD003958-01, A.R.), the Howard Hughes Medical Institute (HHMI), the Klarman Cell Observatory (A.R.), the United States-Israel Binational Science Foundation (BSF) (N.F.), the European Research Council (ERC) (309788, I.A.; 340712 N.F.), the Israel Science Foundation (ISF) (1782/11, I.A.), I-CORE (I.A. and N.F.), and the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (NIEHS) (P.J.B.).

Received: May 21, 2014

Revised: September 9, 2014

Accepted: October 22, 2014

Published: December 11, 2014

REFERENCES

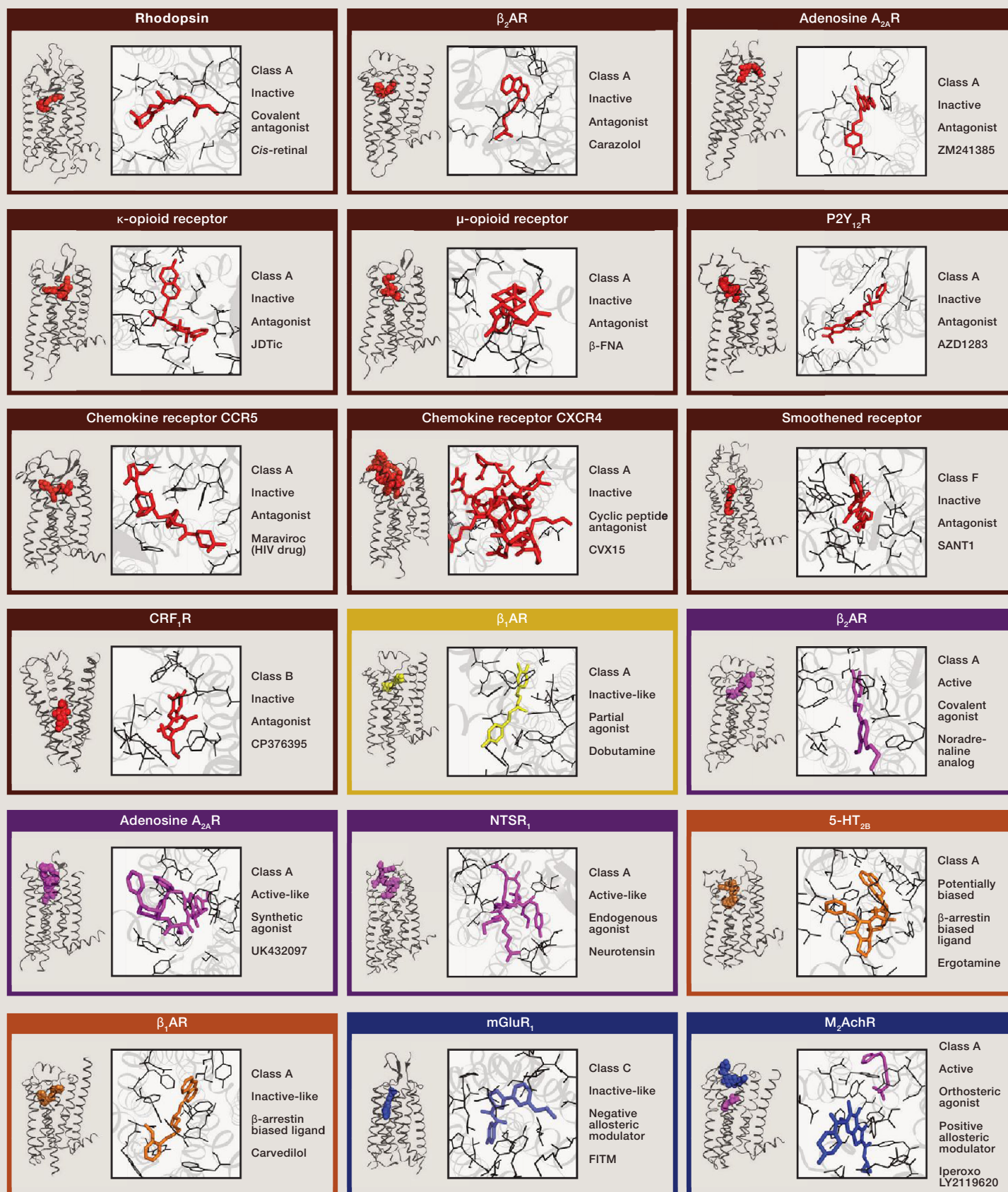
- Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D.S., Busby, M.A., Berlin, A.M., Sivachenko, A., Thompson, D.A., Wysoker, A., Fennell, T., et al. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat. Methods* 10, 623–629.
- Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O., et al. (2009). Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* 326, 257–263.
- Audibert, A., Weil, D., and Dautry, F. (2002). In vivo kinetics of mRNA splicing and transport in mammalian cells. *Mol. Cell. Biol.* 22, 6706–6718.
- Brooks, S.A., and Blackshear, P.J. (2013). Tristetraprolin (TTP): interactions with mRNA and proteins, and current thoughts on mechanisms of action. *Biochim. Biophys. Acta* 1829, 666–679.
- Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927.
- Carballo, E., Lai, W.S., and Blackshear, P.J. (1998). Feedback inhibition of macrophage tumor necrosis factor- α production by tristetraprolin. *Science* 281, 1001–1005.
- Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al.; FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) (2005). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563.
- Carpenter, S., Aiello, D., Atianand, M.K., Ricci, E.P., Gandhi, P., Hall, L.L., Byron, M., Monks, B., Henry-Bezy, M., Lawrence, J.B., et al. (2013). A long non-coding RNA mediates both activation and repression of immune response genes. *Science* 341, 789–792.
- Chew, G.L., Pauli, A., Rinn, J.L., Regev, A., Schier, A.F., and Valen, E. (2013). Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development* 140, 2828–2834.
- Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368–373.
- Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848.
- Danecek, P., Nellåker, C., McIntyre, R.E., Buendia-Buendia, J.E., Bumpstead, S., Ponting, C.P., Flint, J., Durbin, R., Keane, T.M., and Adams, D.J. (2012). High levels of RNA-editing site conservation amongst 15 laboratory mouse strains. *Genome Biol.* 13, 26.
- Dölken, L., Ruzsics, Z., Rädle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., and Koszinowski, U.H. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* 14, 1959–1972.
- Eser, P., Demel, C., Maier, K.C., Schwalb, B., Pirkli, N., Martin, D.E., Cramer, P., and Tresch, A. (2014). Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. *Mol. Syst. Biol.* 10, 717.
- Fuchs, G., Voicheck, Y., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2014). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol.* 15, R69.
- Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., Robinson, J., Minie, B., Chevrier, N., Itzhaki, Z., et al. (2012). A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol. Cell* 47, 810–822.

- Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312, 75–79.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141.
- Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C., and Glass, C.K. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* 51, 310–325.
- Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* 7, 1009–1015.
- Kleinman, C.L., and Majewski, J. (2012). Comment on “Widespread RNA and DNA sequence differences in the human transcriptome”. *Science* 335, 1302, author reply 1302.
- Lai, W.S., Parker, J.S., Grissom, S.F., Stumpo, D.J., and Blackshear, P.J. (2006). Novel mRNA targets for tristetraprolin (TTP) identified by global analysis of stabilized transcripts in TTP-deficient fibroblasts. *Mol. Cell. Biol.* 26, 9196–9208.
- Lee, M.T., Bonneau, A.R., Takacs, C.M., Bazzini, A.A., DiVito, K.R., Fleming, E.S., and Giraldez, A.J. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* 503, 360–364.
- Lee, M.T., Bonneau, A.R., and Giraldez, A.J. (2014). Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* 30, 581–613.
- Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A., and Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* 7, 709–715.
- Li, J.B., Levanon, E.Y., Yoon, J.K., Aach, J., Xie, B., Leproust, E., Zhang, K., Gao, Y., and Church, G.M. (2009). Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* 324, 1210–1213.
- Li, M., Wang, I.X., Li, Y., Bruzel, A., Richards, A.L., Toung, J.M., and Cheung, V.G. (2011). Widespread RNA and DNA sequence differences in the human transcriptome. *Science* 333, 53–58.
- Lin, W., Piskol, R., Tan, M.H., and Li, J.B. (2012). Comment on “Widespread RNA and DNA sequence differences in the human transcriptome”. *Science* 335, 1302, author reply 1302.
- Neeman, Y., Levanon, E.Y., Jantsch, M.F., and Eisenberg, E. (2006). RNA editing level in the mouse is determined by the genomic repeat repertoire. *RNA* 12, 1802–1809.
- Pandya-Jones, A., Bhatt, D.M., Lin, C.H., Tong, A.J., Smale, S.T., and Black, D.L. (2013). Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA* 19, 811–827.
- Pauli, A., Valen, E., Lin, M.F., Garber, M., Vastenhouw, N.L., Levin, J.Z., Fan, L., Sandelin, A., Rinn, J.L., Regev, A., and Schier, A.F. (2012). Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* 22, 577–591.
- Pickrell, J.K., Gilad, Y., and Pritchard, J.K. (2012). Comment on “Widespread RNA and DNA sequence differences in the human transcriptome”. *Science* 335, 1302, author reply 1302.
- Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat. Biotechnol.* 29, 436–442.
- Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J., et al. (2010). The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.* 38, D613–D619.
- Schier, A.F. (2007). The maternal-zygotic transition: death and birth of RNAs. *Science* 316, 406–407.
- Shalem, O., Dahan, O., Levo, M., Martinez, M.R., Furman, I., Segal, E., and Pilpel, Y. (2008). Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol. Syst. Biol.* 4, 223.
- Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* 16, 1128–1133.
- Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508, 66–71.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476.
- Windhager, L., Bonfert, T., Burger, K., Ruzsics, Z., Krebs, S., Kaufmann, S., Malterer, G., L'Hernault, A., Schilhabel, M., Schreiber, S., et al. (2012). Ultra-short and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Res.* 22, 2031–2042.
- Zeisel, A., Köstler, W.J., Molotski, N., Tsai, J.M., Krauthgamer, R., Jacob-Hirsch, J., Rechavi, G., Soen, Y., Jung, S., Yarden, Y., and Domany, E. (2011). Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Mol. Syst. Biol.* 7, 529.
- Zhang, Y., Vastenhouw, N.L., Feng, J., Fu, K., Wang, C., Ge, Y., Pauli, A., van Hummelen, P., Schier, A.F., and Liu, X.S. (2014). Canonical nucleosome organization at promoters forms during genome activation. *Genome Res.* 24, 260–266.

SnapShot: GPCR-Ligand Interactions Cell

Eshan Ghosh, Kumari Nidhi, and Arun K. Shukla

Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur,
208016, India





Abstract
submission
deadline:

January 16
2015

Early bird
deadline:

March 27
2015

This symposium will bring academic and industry researchers and clinicians together to explore the many facets of cancer immunology and discuss how tumor-associated inflammation can be reprogrammed for therapeutic benefit.

Keynote Speakers

James Allison, USA
Lisa Coussens, USA

Speakers

Rafi Ahmed, USA
Vincenzo Bronte, ITA
George Coukos, SUI
Mikala Egeblad, USA
Jerome Galon, FRA
Johanna Joyce, USA
Michael Karin, USA
Alberto Mantovani (co-organizer), ITA
Ignancio Melero, ESP

Ira Mellman (co-organizer), USA
Peter Murray, USA
Jeffrey Pollard, UK
Fiona Powrie, UK
Carola Ries, GER
Alexander Rudensky, USA
Michel Sadelain, USA
Giorgio Trinchieri, USA
Laurence Zitvogel, FRA

Organizers

Alberto Mantovani, *University of Milan, ITA*
Ira Mellman, *Genentech, USA*
Elizabeth Thompson, *Scientific Editor, Immunity*

cell-symposia-cancerandinflammation.com

cell.com/symposia